370
7.9.70

Psychological Review

THEODORE M. NEWCOMB, Editor
University of Michigan

Consulting Editors

SOLOMON E. ASCH
ROBERT R. BLAKE
STUART W. COOK
CLYDE H. COOMBS
WILLIAM K. ESTES
LEON FESTINGER
W. R. GARNER
JAMES J. GIBSON
D. O. HEBB
HARRY HELSON
E. R. HILGARD
CARL I. HOVLAND
E. LOWELL KELLY
DAVID KRECH
ROBERT W. LEEPER
KENNETH MACCORQUODALE
ROBERT B. MACLEOD
DAVID C. MCCLELLAND
GEORGE A. MILLER
FREDERICK MOSTELLER
GARDNER MURPHY
OSCAR OESER
CARL PFAFFMANN
CARROLL C. PRATT
JOHN P. SEWARD
DAVID SHAKOW
RICHARD L. SOLOMON
ELIOT STELLAR
S. S. STEVENS
ERIC TRIST
EDWARD L. WALKER
ROBERT W. WHITE

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
HERBERT NELL, Editorial Assistant

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

VOLUME 65, 1958

Bureau Ednl. Psy. Researc
DAVID HARE TRAINING COLLEG
Dated ....
Accs. No ....

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Bureau Ednl. Research
DAVID HA NG CO E
Dated 7.9.70
Acc. No. J-370

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 65

Anastasi, Anne. Heredity, Environment, and the Question "How?" 197
Arbit, J. Avoidance Conditioning Through Irradiation: A Note on Physiological Mechanisms and Psychological Implications 167
Bevan, W. Perception: Evolution of a Concept 34
Bovard, E. W. The Effects of Early Handling on Viability of the Albino Rat 257
Brown, R. How Shall a Thing Be Called? 14
Davis, R. C. The Domain of Homeostasis 8
Edwards, W. *Luce*, R. D.
Garner, W. R. Symmetric Uncertainty Analysis and Its Implications for Psychology 183
Gilbert, T. F. Fundamental Dimensional Properties of the Operant 272
Graham, C. H. Sensation and Perception in an Objective Psychology 65
Green, E. J. A Simplified Model for Stimulus Discrimination 56
Hollander, E. P. Conformity, Status, and Idiosyncrasy Credit 117
Jessor, R. The Problem of Reductionism in Psychology 170
Jones, H. G. The Status of Inhibition in Hull's System: A Theoretical Revision .. 179
Kagan, J. The Concept of Identification 296
Keehn, J. D. *See* Mowrer, O. H.
Luce, R. D., and Edwards, W. The Derivation of Subjective Scales from Just Noticeable Differences 222
Meissner, W. W. Nonconstructural Aspects of Psychological Constructs 143
Mowrer, O. H., and Keehn, J. D. How Are Intertrial "Avoidance" Responses Reinforced? 209
Murdock, B. B., Jr. Intralist Generalization in Paired-Associate Learning 306
Newell, A., Shaw, J. C., and Simon, H. A. Elements of a Theory of Human Problem Solving 151
O'Neil, W. M. Basic Issues in Perceptual Theory 348
Peak, Helen. Psychological Structure and Psychological Activity 325
Pratt, C. C. Herbert Sidney Langfeld: 1879–1958 321
Restle, F. Toward a Quantitative Description of Learning Set Data 77
Rhine, R. J. A Concept-Formation Approach to Attitude Acquisition 362
Riley, D. A. The Nature of the Effective Stimulus in Animal Discrimination Learning: Transposition Reconsidered 1
Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain 386
Rozeboom, W. W. "What Is Learned?"—An Empirical Enigma 22
Shaw, J. C. *See* Newell, A.
Shepard, R. N. Stimulus and Response Generalization: Deduction of the Generalization Gradient from a Trace Model 242
Simon, H. A. *See* Newell, A.
Slack, C. W. Trial Design in Human Experiments 92
Walker, E. L. Action Decrement and Its Relation to Learning 129
Wallach, M. A. On Psychological Similarity 103
Wegner, Norma, and Zeaman, D. Strength of Cardiac Conditioned Responses with Varying Unconditioned Stimulus Durations 238
Wohlwill, J. F. The Definition and Analysis of Perceptual Learning 283
Vanderplas, J. M. Transfer of Training and Its Relation to Perceptual Learning and Recognition 375
von Fieandt, K. Toward a Unitary Theory of Perception 315
Yates, A. J. Symptoms and Symptom Substitution 371
Zeaman, D. *See* Wegner, Norma.

BUREAU OF EDNL. & PSY. RESEARCH LIBRARY

VOL. 65, NO. 1 JANUARY, 1958

THE PSYCHOLOGICAL REVIEW

THE NATURE OF THE EFFECTIVE STIMULUS IN ANIMAL DISCRIMINATION LEARNING: TRANSPOSITION RECONSIDERED

DONALD A. RILEY

University of California

When an animal learns to discriminate between two stimuli lying on the same physical dimension, what is the nature of the effective stimulus with which the response is associated?

In 1937, Spence (13) presented a theory which assumed that the animal learns to approach the absolute properties of the positive stimulus and that independent of this, it learns not to approach the absolute properties of the negative stimulus. In those situations in which the animal learns to discriminate between two points on the same physical dimension, Spence defined the effective stimulus in conventional physical units—for example, the number of square centimeters in a size discrimination, or the number of foot candles in a brightness discrimination. Spence contrasted this absolute definition of the effective stimulus with Köhler's earlier assumption (7) that the animal responds to the relationship between the two stimuli, and learns to go to the larger or the brighter of the two stimuli. Spence made some additional assumptions that enabled him to predict the transposition of the learned response to pairs of test stimuli which differ from the training stimuli. In contrast to the relational theory, which made no quantitative predictions, Spence's theory predicted that transposition would decrease as the test stimuli became less and less similar to the training stimuli. This prediction has been confirmed in several experiments (1, 6, 9, 13).

This paper will present evidence that Spence's assumption that the animal responds to the absolute properties of the stimulus, defined in units such as absolute size or luminance, is untenable. This evidence also casts doubt on Spence's explanation of the decline in transposition. Two alternative reinterpretations of the facts of transposition will then be considered. Before presenting the new evidence, Spence's theory will be described, together with an example of the type of evidence supporting it.

Near and far transposition tests. The meaning of the terms "near transposition test" and "far transposition test" can be given schematically. Let letters A through J represent points on a physical continuum, such that the ratio of B to A is the same as C to B, and so on. If the *S* is originally trained to choose B in the presence of A and B, then a test with B and C is a near transposition test. A test with any other pair of adjacent stimuli (e.g., H and I) is a

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE

far transposition test. For example, if in original training the ratio of the magnitude of the positive to the negative stimulus is 2/1, then in the near transposition test, the ratio will be 4/2. In the far transposition test, the ratio might be 64/32, or 128/64. Although Spence and others have discussed other combinations of stimuli, in the interest of simplicity and brevity, the discussion here will be restricted to the case in which the test stimuli are of greater magnitude than the training stimuli.

The sample experiment which will now be described in this simplified form is one by Kendler (6). Two stimuli serving as cues to the presence or absence of reward were presented to the *S*s (in this case, rats). These stimuli differed only in that one had a photometric value of .011 foot candle while the other had a value of .63 foot candle. Going to one of the stimuli (e.g., the brighter) was always followed by reward. Going to the other was followed by nonreward. After learning, the animal was tested with other pairs of stimuli. Different animals were tested at different intensities, but in each test the ratio between the two stimuli was the same as in original training. In each test, the animals preferred the more intense stimulus, but the magnitude of the preference decreased as the intensity of the test pairs increased.[1] That is, far transposition is weaker than near transposition. Similar results have been found with size as the critical dimension (1, 9, 13).

The absolute theory. Because Spence's theory is given in detail in many references, it will be summarized here only in very brief form. It assumes that if the animal is regularly rewarded when it approaches one point on a stimulus dimension, and is regularly nonrewarded when it approaches another point on the same dimension, the animal will develop corresponding approach (Excitatory) and nonapproach (Inhibitory) tendencies. These E and I tendencies will be elicited by other points on the dimension, but the intensity of the elicited tendency will decline as the distance of the test point from the training point increases. The shapes of the E and I gradients are positively decelerated, and, in the simplest case, have the same equation. The tendency to make an approach response to any point on the dimension will be a simple function of the difference between the two opposing tendencies. Since, at points above the training stimuli on the stimulus continuum, the I generalization gradient will fall more rapidly than the E gradient, transposition will occur. As both curves fall, the greater E-minus-I value of the more intense stimulus becomes relatively smaller, and less transposition occurs.

Limitations in the generality of the absolute theory. While it is true that this absolute theory does account for transposition, and for the drop in transposition in the far test, some recent research, especially that of Lawrence and his associates, appears to present findings incompatible with Spence's account of transposition. Baker and Lawrence (2) found transposition with rats following conventional training on a size-discrimination problem, but did not find transposition when the training stimuli were presented successively. Unfortunately, the two conditions differed in one respect other than simultaneity vs. successiveness. The animals in the simultaneous condition learned to criterion, and thus received almost 100% reinforcement toward the end of training. In contrast, the animals in the successive condition had the same number of trials with the negative stimulus as with the positive stimulus throughout training, and consequently were reinforced only 50% of the time. Baker and Lawrence pointed out that in spite of this methodological flaw, the results are not consistent with Spence's theory.

---

[1] The amount of transposition in Kendler's experiment did not decline monotonically when the test stimuli were more intense than the training stimuli. The animals transposed more to the most remote stimuli than to those of intermediate value. The reason for this is not clear, but one possibility is that the voltage changes that were introduced changed the wave length of the light. This in turn might influence brightness discrimination.

Another apparent contradiction of the absolute theory is an experiment by Lawrence and DeRivera (10). Animals were trained in a discrimination problem in which, on each trial, two doors showed the same stimulus pattern. The bottom half of each door was always the same medium gray (shade No. 4 in a set of seven shades). The top half of each was either a shade of gray three steps lighter (No. 1) than the bottom half, or a shade of gray three steps darker (No. 7) than the bottom. The animals were trained to jump to one door when the upper parts were lighter than the lower, and to jump to the other door when the upper parts were darker. After training, the animals were given combinations that tested between a relational interpretation of learning and an absolute interpretation. For example, with No. 3 over No. 1, the animal should jump to the side associated with the lighter gray, according to the absolute hypothesis, because both cards are lighter than No. 4; but since the relationship is dark over light, they should jump the other way according to the relational hypothesis. The results predominantly supported the relational hypothesis.

Clearly, the results of Lawrence *et al.* appear to be in contradiction to Spence's theory that the animal responds to the cues in an independent fashion. As a possible solution of these apparent contradictions, Lawrence and DeRivera (10) and Hilgard (5) have suggested that under some circumstances the animal responds relationally, and that under other circumstances it responds absolutely. That is, perhaps it responds absolutely in a simple simultaneous discrimination situation, but not in others. It is possible, however, that *the decline in transposition in the far test* has not yet been adequately explained, and that an understanding of *this* phenomenon will resolve the apparent inconsistencies. Then, we will be in a better position to state testable alternatives concerning the nature of transposition.

A contrast interpretation of the drop in transposition. If, instead of defining each stimulus merely in terms of the patch alone, it is defined as a specific relationship between the value of the patch and the value of the surround, then the decline in transposition might result merely from the change in the relationship being responded to. Such a proposal is, of course, not new. In human perception, the effect of contrast on apparent brightness has been known for many years. Recently, Wallach (15) has shown that two test stimuli of radically different illumination will be judged approximately equal in brightness if the ratio of the test-stimulus intensity to surrounding intensity is the same for both stimuli. Similarly, Gibson (3) has shown that perceptual size constancy in humans depends on such factors as the texture gradient of the stimulus field surrounding the test stimulus. It is not unreasonable to assume that apes and rats might behave in a similar way.

An experimental test of the contrast hypothesis. The specific hypothesis to be tested is that after a brightness discrimination has been learned, there will be a drop in transposition on the far test only if the ratios of the test stimuli to their surround are changed.

The experiment was similar to Kendler's in that the rats were trained to discriminate brightness. All animals were trained in a modified Lashley jumping apparatus to jump to the more intense of two stimuli. The doors and the surrounding stimulation were illuminated from projectors hung near the top of the room. The different illuminations on the doors and the surround were determined solely by slides in the projectors, and by the size of openings in iris diaphragms on the projectors. During training, the brightnesses of the two doors and the surround, as measured by a Zeiss photometer, were in the ratio of 10/3/1, the surround having the value of 1. The surround had a reflected illumination of .05 foot candle.[2] During training, the rats were run 10 trials a day by the noncorrection method. Forced trials were used to break position habits. Following training to a criterion of 19/20 correct responses, all animals were tested under four different conditions. Each animal was tested on only one condition in a day, and on days between tests the animal was retrained to a criterion of 9/10 correct responses. Each test consisted of ten trials during which the animal was rewarded regardless of the correctness of its choice. The order of conditions was counterbalanced. The four conditions were: (*a*) near absolute—the two test patches were transposed one step up, and the surrounding stimulation was the same as in original training; (*b*) far absolute—the two test patches were made 5 steps brighter and the surrounding stimulation was the same as in original training; (*c*) near ratio—the same as (*a*) except that the illumination of the surround was increased so that the ratio of intensities of the test patches to the surround was the same as in original training; and (*d*) far ratio—the same as (*b*) except that the illumination of the surround was increased so that the ratio of the intensities of the test patches to the surround was the same as in original training. Sixteen hooded rats of the Long-Evans strain were used.

Predictions. For this experiment to offer a test between the absolute and the contrast hypotheses, it is essential that there be significantly less transposition on the far absolute condition than on the near absolute condition. If this requirement is met, then the absolute theory predicts a similar drop from the near ratio to the far ratio condition, for the only two parts of the stimulus situation that have been differentially reinforced are being changed, just as they are in the absolute conditions. On the other hand, the contrast hypothesis predicts a drop in transposition from the near to the far absolute condition but no such drop in the ratio condition. If the ratio present in training is present in the test, then the animal should transpose, regardless of the absolute values of the stimuli.

TABLE 1

MEAN NUMBER OF RESPONSES TO THE BRIGHTER STIMULUS IN TEN CHOICES

	Near Absolute	Far Absolute	Near Ratio	Far Ratio
Mean	9.25	6.94	9.19	9.69
SD	1.09	1.84	1.36	.54

Results. The main results are shown in Table 1. Their meaning is unambiguous. The animals transpose almost perfectly in all except the far absolute condition. There were no effects of order. The only effect is a highly significant drop in transposition under the far absolute condition. A *t* test was run in which for each *S* the difference between near and far ratio was subtracted from the difference between near and far absolute. The *t* of 4.87 is significant at the .0002 level of confidence, for 15 degrees of freedom.

Interpretation of the results. It is clear, at least in the case of brightness discrimination, that the animal responds to the relationship between the test stimulus and the background. It is a change in this relationship rather than a change in the absolute value of the stimulus that causes the breakdown in transposition. While this experiment does not provide an explanation of transposition per se, it does show that the drop in transposition in the far test is not critical evidence for Spence's explanation of transposition. The most immediate effect of this finding is to force a reconsideration of possible interpretations of transposition. Two such possible interpretations that seem to this writer to be reasonable alternatives will be considered. Again, for simplicity, the discussion will use brightness discrimination as the example.

---

[2] I wish to thank J. Gulberg, Department of Zoology, University of California, for supervising the measurement of the physical stimuli, and Arnold Mechanic, who participated in the running of the experiment.

A reformulation of the E and I gradients hypothesis. If the effective stimulus is redefined as the ratio of the training or test patch to the surround, then it is possible to maintain the assumption that the animal responds separately to the two training stimuli. In this case, the appropriate description of the stimulus intensity dimension will also be expressed in such ratios. When an animal has been trained to approach one stimulus (defined in ratio terms) and not to approach the other, E and I tendencies will be reduced to the degree that test ratios change from training ratios. If these redefined generalization gradients are of the form postulated by Spence, this hypothesis can then account for transposition and for the decline in transposition with changing stimulus ratios.

Can this interpretation account for the findings of the other experiments that were critical of Spence's absolute hypothesis? Lawrence and DeRivera's experiment is somewhat complex for the present analysis, because relationships within the stimulus patch were changed in the tests, but the surrounding stimulation was always the same as in training. Nevertheless, their results at least do not appear inconsistent with this hypothesis. If the two parts of the stimulus patch are responded to in the way that the stimulus surround was responded to in the present experiment, then their animals learned to jump left when the ratio on the cards was in one direction and right when it was in the other direction. As the ratios were changed in the test conditions, the tendency to make the relational response decreased.

On the other hand, this reformulation of the absolute theory does not account for the failure of Baker and Lawrence to find transposition following successive training. Their experiment therefore seems critical, and should be repeated with the same pattern of reinforcements in the successive and simultaneous groups.

A reformulation of the relational hypothesis. Since the decline in transposition in the far test cannot be considered as evidence for an absolute definition of the stimulus, the question of the adequacy of Köhler's original relational hypothesis is reopened. In his discussion of transposition, Köhler merely described the animal as having learned to go to the larger or lighter part of the stimulus. This restricted theory is still inadequate, for the reason pointed out by Spence. It cannot account for the decline in transposition in the far test. Nor does it help Köhler's explanation to say that the figure-ground relationships determine the phenomenal brightness of the two test stimuli, and that the animal approaches the stimulus which is phenomenally brighter, for this still would not predict a breakdown in transposition in the far test. Another unworkable relational assumption is that the animal responds to the ratio between the two figure-ground ratios. This hypothesis would state that in the present experiment the animal responds to the relationship of 10/1 (the brighter) to 3/1 (the dimmer)—that is, to a relationship of 10/3 or 3.33. Now in the critical far-absolute test, the values are about 4,100/1 for the brighter stimulus and 1,230/1 for the dimmer. This of course reduces to 3.33 and perfect transposition is predicted. Since transposition is known to decline under these circumstances, this hypothesis is inadequate.

It seems clear that a workable relational hypothesis will be one that predicts perfect transposition as long as all stimulus values in the critical dimension are multiplied by the same factor. If the values in the original training are in the ratio of 10/3/1, then any test situation in which the ratio is the same must give perfect transposition.[3] To the extent that any of the three parts of the stimulus deviates from this, transposition will decline. One possible interpretation of such behavior would be that the animal approaches a stimulus that is defined by two intensity ratios, and does not approach one that is defined by two other intensity ratios. That is, the animal learns to approach a training stimulus which has an intensity that is a certain multiple of the intensity of the surround and a certain multiple of the intensity of the other training stimulus. Similarly, the animal learns not to approach a stimulus that has an intensity that is a certain multiple of the intensity of the surround and a certain fraction of the intensity of the other training stimulus. This hypothesis has no trouble in accounting for a decline in transposition in the far-absolute situation of the present experiment. It is not so clear why there is no drop in transposition in the near-absolute situation, for here there is also a change in the ratios. One *ad hoc* interpretation is that, under the usual conditions of the transposition experiment, the near-absolute test is not a discriminably different relationship from the relationship present during training. There are a variety of reasons why this might be so. For example, since the difference between the training stimuli and the surround is substantial, a change of only one step may be relatively hard to detect. Further, twenty-four hours intervene between the last training trial and the first test trial. This increases the difficulty of discriminating between the two relationships.

This particular relational hypothesis is very similar to the reformulated absolute hypothesis. It differs only in that each critical stimulus is defined by two brightness gradients rather than just one. A direct test between these two hypotheses can be made in an experiment in which the stimulus area surrounding each stimulus patch may be independently varied. Then, following training identical to the present experiment, the ratio of test stimulus to surround could be kept constant, as in the ratio conditions of the present experiment, but the relationship between the two test stimuli could be changed from that of the training situation; e.g., the stimulus-surround value of the positive stimulus might be kept at 10/1, while the negative is changed from a training value of 3/1 to a test value of 30/10. As long as the stimulus-surround ratios are constant, the reformulated absolute hypothesis should predict that the animal would continue to approach the test stimulus that has the same ratio as the positive training stimulus, even though the relationship between the test stimuli is changed. This should be true regardless of the amount or direction of the change. The proposed relational hypothesis would predict breakdown in transposition as the ratio of the test stimuli to each other becomes more different from that of the training stimuli, even though the stimulus-surround ratios are the same as in original training.

There may be other relational hypotheses that would be consistent with the present facts. For example, Teas and Bitterman (14) have described conditions under which the rat appears to respond to the stimulus as a total configuration rather than responding selectively to different components of the stimulus. The only difficulty in applying this hypothesis to the present data is that this experiment was run under conditions that are probably closer to the condition in their study that elicited selective reactions from the rat.

---

[3] Although Köhler explicitly made this point when discussing the general problem of transposition (cf. 8, p. 218), he did not apply this principle either in the design or in the analysis of specific experiments. In discussing the experiments, only the changes in the test patches were considered.

Two further implications of the present experiment. It is worth noting that the present study has implications not only for transposition, but for the closely related problem of stimulus generalization along quantitative continua. It appears that there is a real possibility that if the relationships existing in original conditioning are present in generalization tests, stimulus generalization will be complete regardless of absolute stimulus values. Directly related to this point is an experiment by Perkins (12) that indicates that stimulus intensity dynamism is probably a contrast phenomenon.

The second point is more of a question than a direct implication of the present experiment. How much of transposition might be accounted for by subcortical or even retinal mediation rather than by cortical processes? Some time ago, Hebb (4) presented evidence that brightness discrimination is probably controlled subcortically. More recently, Motokowa (11) has presented evidence from retinal recordings of the frog indicating that action potentials are related to brightness contrast rather than intensity.

SUMMARY

Evidence has been presented that in animal discrimination learning, as in human perception, the effective stimulus to which the animal responds must be described as a relationship between parts of a stimulus complex. This finding allows a reinterpretation of Spence's explanation of the decline in transposition when test stimuli are remote from training stimuli. It also sets the conditions for a reinterpretation of transposition. Two alternatives are discussed.

REFERENCES

1. ALBERTS, E., & EHRENFREUND, D. Transposition in children as a function of age. *J. exp. Psychol.*, 1951, **41**, 30–38.
2. BAKER, R. A., & LAWRENCE, D. H. The differential effects of simultaneous and successive stimuli presentation on transposition. *J. comp. physiol. Psychol.*, 1951, **44**, 378–382.
3. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.
4. HEBB, D. O. The innate organization of visual activity. III. Discrimination of brightness after removal of the striate cortex in the rat. *J. comp. Psychol.*, 1938, **25**, 427–437.
5. HILGARD, E. R. *Theories of learning.* (2nd ed.) New York: Appleton-Century-Crofts, 1956.
6. KENDLER, T. S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, **40**, 552–562.
7. KÖHLER, W. Nachweis einfacher strukturfunktionen beim schimpansen und beim haushuhn: Über eine neue methode zur untersuchung des bunten farbensystems. *Abh. Preuss. Akad. Wissen,* 1918, Physikalisch-Mathematische Klass., No. 2, 1–101.
8. KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1929.
9. KUENNE, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
10. LAWRENCE, D. H., & DERIVERA, J. Evidence for relational discrimination. *J. comp. physiol. Psychol.*, 1954, **47**, 465–471.
11. MOTOKAWA, K. Physiological induction in human retina as basis of color and brightness contrast. *J. Neurophysiol.*, 1949, **12**, 475–488.
12. PERKINS, C. C. The relationship between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.
13. SPENCE, K. W. The differential response in animals to stimuli varying in a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
14. TEAS, D. C., & BITTERMAN, M. E. Perceptual organization in the rat. *Psychol. Rev.*, 1952, **59**, 130–140.
15. WALLACH, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310–324.

(Received June 17, 1957)

Psychological Review
Vol. 65, No. 1, 1958

THE DOMAIN OF HOMEOSTASIS[1]

R. C. DAVIS

Indiana University

In the halls of our neighboring physiology department hangs the celebrated motto of Claude Bernard: *Le fixité du milieu interieur est la condition de la vie libre.* Of late, in our own discipline, as well as in physiology, one may observe a spreading and unchallenged faith in the virtue of this apothegm and its sequel, that all the "vital mechanisms" are devoted to the maintenance of this stability. It is thought to contain the germ which, cultivated by Cannon (3) into the full-flowered doctrine of homeostasis, now offers general shelter for psychologists. The "maintenance of equilibrium" has often been taken as the unifying principle of "motivation," and more recently as fundamental to all action of the living (1, 5).[2] As a model for behavior the principle of homeostasis presents itself in rivalry with the time-honored formula of stimulus-response or of any more modern equivalent.

Clearly this is a possibility of such importance that the meaning and domain of "homeostasis" should be given careful thought. Homeostasis and adaptation are sometimes treated as though one implied the other. But definitions of the two terms are usually not thus restricted, and the facts do not justify an inference of this sort. *The Wisdom of the Body* in both title and content shows that Cannon was strongly impressed with the "adaptiveness" of homeostatic regulation, and he showed no qualms about including purposive clauses in his sentences. He observed, nevertheless, that ". . . at times effects are produced which apparently have little or no value to the organism," and that infants and lower animals lack many homeostatic mechanisms. (Though he did not discuss the point, it is doubtful that he meant these were ill adapted.) The meaning of "adaptation" is none too clear in a scientific context, and even if it were established that a certain event is adaptive, it would follow only with an unknown probability that it is also homeostatic.

Two definitions of homeostasis seem possible on the basis of current and historical usage. It is difficult to say which is more legitimate, for they were doubtless commingled in the original writings. Cannon, and in the earlier period Bernard, were evidently led to the general conception because of some physiological facts which came to light. These discoveries showed that certain physiological variables are more constant than one would expect in a changing world, and that there are some specific mechanisms, now called negative feedbacks, which act, when a change occurs, in such a manner as to minimize the change. Either the first or the second part of the preceding sentence may be taken as the basic meaning of homeostasis.[3] In other words, homeostasis can mean, and evidently does sometimes mean, simply the existence of a certain constancy in the face of change. On the other hand, one may use the name only for a certain type of mechanism. In the one case it is an end, in the other a means. (Of course, both meanings at once may be intended, though such restriction seems unusual.)

---

[1] Based on a lecture given at McGill University and University of Oslo, September and October, 1956. Assistance in preparation was given by Contract Nonr 908-03, Office of Naval Research.

[2] The form of Bernard's principle implies that he did not think of it as universal; the "free life" seems to be something beyond the regulation he is describing. It is very clear that Cannon (3) did not think homeostasis ruled everything, for he asks ". . . freedom for what?" and answers "In summary, then, we find the organism liberated for its more complicated and socially important tasks because it lives in a fluid matrix, which is automatically kept in a constant condition." Generally he regards the action of skeletal muscles as disturbing to homeostasis.

Though often not distinguished, the two meanings are not the same. Stability, resistance to change, or even restoration of a prior state is not invariably a sign of a negative feedback. This could, no doubt, be demonstrated by a mathematical argument, but the point can be made as convincingly with less formality.

Evidently the behaving organism is one of those physical structures called "systems." It qualifies as one of these because it has outputs—movements of body, hands, vocal cords, etc.—in some kind of relation to inputs, some of which are usually called stimuli and some of which are more general conditions. If we follow the usual presumption, this relation results from the operation of the same physical laws which apply to other systems. Accepting this identification, whatever we may know of systems—from everyday experience as well as from the mathematically developed theory of systems—will come to our aid.

Although one could not hope to decide exactly what happens on such general grounds, the field of possibilities can be helpfully delimited. If, for example, the principle of conservation of energy places certain restrictions on a system, it would be an extraordinarily bold hypothesis which would propose that the organism transcends them.

In the case at hand—i.e., the organism's stability—we may look at simpler systems to find, if we can, the same sort of stability. To begin with a very humble example, let us reflect on the behavior of an ordinary screen door being pushed by a small boy who likes to experiment. This, too, is a system with an input. If he unhooks the spring the boy will be pleased to find that the more push he gives the farther the door will go before it stops. It will be even more of a delight to find, after the spring is attached, that the door now pushes back at him. The more he pushes, the more the door reacts, and when he lets go the door will return to its original place. In more technical words, he has now observed that the door has a nonlinear response to a step-function input, and returns to zero after a square-wave input. It is displaced less than expected by an input force, and has "automatic" stabilization at a point of equilibrium. By other examples one could show that such a system can oscillate (as even a screen door can bang twice if conditions are right). These kinds of behavior are in fact to be found in any system which contains a reacting element—such an element as produces a restoring force as a funtion of input force, and this is a very common feature of systems. Restoring forces are provided by springs or other elastic bodies, condensers and inductors, heat-absorbing elements and chemical processes. Some elements of this kind are present in all systems, and it is a major engineering problem to construct a system in which their effect is too small to matter. With a moment's thought we can find all of these elements, except inductors, in the living organism. There is elasticity in all the tissues, and in muscles and blood vessels it is known to be important; electrical and heat capacities exist throughout the body, and the piling up of chemical energy occurs in many places.

---

[3] Cannon wrote ". . . constancy is in itself evidence that agencies are acting, or ready to act to maintain this constancy." This implication would cover other mechanisms than feedback, and indeed he considers "storage" of simple capacitative sort an example of such an agency. In considering specific constancies, however, Cannon apparently takes it for granted that the "agency" is a feedback, even though the mechanism is unknown (e.g., 3, Ch. 8).

Clearly then, the "stability" of a response system is not much of an argument for the existence of a feedback mechanism. The same effects, to the closest approximation we are likely to be able to measure, can be brought about in a simpler fashion with a system which is reactive. To complete the indirect argument one would need to exclude in some way the possibility of the simpler mechanism in the case at hand; still better evidence would come from a direct demonstration of negative feedback.

At this point it is interesting to note a statistical argument which leads to the same conclusion.[4] Suppose a system can be in a certain number of states, ordered in some respect from lowest to highest. For simplicity, then suppose that each of these states is equally probable and in no way contingent upon any other. Upon finding the system in a state toward the top of the scale one would predict that its next state, if there is one, will be a lower one, simply because there are more states below than above it in the scale. Their summed probabilities will be greater. Certainly if the system is already at the top of the scale, its next state will always be lower, because by definition there is nowhere to go but down, and the best prediction would be the mean of the possible states. If the organism is already in its middle state one would, of course, predict its next state as a departure from that, but the mean of the possible states would still be the best prediction. (If this is not to be counted as a "next state," the best prediction would be as close as possible to the mean.) Further, suppose that a measured property, such as blood pressure, is the resultant of a considerable number of factors acting according to this description. The measured variable then tends actually to assume the mean position, merely by the random action of the many factors. An additional property of the model, in either simple or multiple form, is that the probability of finding it in a state deviating from the mean by a certain value decreases with the size of the deviation. (These consequences will also follow from assumptions more general than the ones used for illustration.)

It is hard to believe that the organism is so disjointed as these postulates would make it. Yet one may wonder whether the laws connecting states of the organism are so many that the effect is the same after all.

Inasmuch as stability and equilibration are so common in systems, the term *homeostasis* would diffuse over the landscape if it were used in a functional sense to specify a certain result. If it settles evenly on everything and distinguishes nothing, then to say that a system is homeostatic is tautological, though possibly useful.

According to its second meaning, homeostasis would be defined as a particular sort of process, regardless of result. Identified as a negative feedback its basic requirement is that some energy is taken from a later part of the system and introduced into an earlier part in such a way as to oppose the change produced there by the original input. Certainly quite a number of such mechanisms in the organism are known; there are others sometimes alleged in rather easy-going interpretation, for which one might desire better credentials. Physiologists (as well as psychologists) seem sometimes to start their interpretations with the postulate that every process is homeostatic.

---

[4] This line of thought was given me by C. J. Burke.

If homeostasis as a process is to serve as a general model for behavior, we should be able to say what information there needs to be to decide that a system is homeostatic. From the definition it follows that three things are necessary: (*a*) We should know what the system is, at least to the extent of identifying its input and output. (*b*) We should be able to demonstrate that energy ("signal") is, or at least can be, carried in a reverse direction. (*c*) We should know that this energy has a negative sign with respect to the original process when it re-enters the main path of the system. If these three things are the meaning of homeostasis in the narrower sense, it is plain that we need a clear idea of the first and a reasonable presumption of the other two, lest an assertion of homeostasis be but empty words. One should be on guard against theorizing which starts from one of the three requirements and draws upon nothing but a rich vocabulary for the other two.

We may contrast two examples of supposed homeostatic mechanisms, one which seems complete and another which seems filled out *ad hoc*. First, in the case of the stretch or myotatic reflex, we know the system concerned. Its output can be taken as the length of the muscle. The feedback is initiated by the receptors in the muscle and is carried backward by the afferents. Its effect is negative with respect to muscle lengthening. In the well known regulation of the heart the case is much less clear. The known fact is that receptors in the large arteries are stimulated by blood pressure increments and initiate a reflex which slows the heart rate. Then, to fill out the homeostatic picture, it is said that this reduces blood pressure. In this instance we collide with some troublesome facts. By Starling's law, an increased rate would simply reduce the stroke volume and have just about no effect on pressure in the arteries. And according to Marey's law, another empirical classic, a slow pulse is associated with high rather than low pressure. To add to the confusion, there are receptors in large veins, responsive to raised pressure there, which *increase* the heart rate by the so-called Bainbridge reflex, and it has been asserted that this too is homeostatic: it keeps down the stroke volume of the heart. At this point it seems we might conclude that the beautiful simplicity of homeostatic regulation has been too appealing, even in physiology.

In this instance the trouble seems to be an uncertainty about what is being regulated by a known mechanism. Of course, so long as a process has any effect at all, it is always possible to find something or invent something which it regulates, and in each case one would have quite a wide choice. This is a particular danger when homeostasis is taken as a general model. We may be offered variables or entities which exist only to be opposed by homeostasis, and conversely homeostatic processes which are said to exist only because they are opposite to some known effect, according to a sort of Hegelian logic.

An extreme example of such speculation, once offered in all seriousness, runs as follows: People have a homeostatic mechanism for maintaining social esteem, or the awareness of it. So, when they become aware that they have fallen in the opinion of their fellow men, they take action to raise their standing until their awareness of esteem reaches the point of equilibrium. The factual basis of this, so far as visible, is that people sometimes do things that others approve of, and it is doubtful that we know any more about this after hearing the theory.

A process entitled to the name of homeostasis by meeting the three criteria would not necessarily return the organism to its undisturbed state. Being a negative feedback it would tend to lessen departure from such a condition, but need not be so powerful as to prevent it. The milder sort of feedback would tend to "overshooting" and oscillation, while admitting some change. Quite possibly this sort of feedback is the more usual in bodily processes.

Where we do have homeostasis, in the sense of a negative feedback, there are implications which would be important to a general theory. One of these deserves the emphasis which may be given it by a special name: *heterostasis*. The following example shows what the word is intended to mean. In humans and some other animals body temperature is said, quite correctly, to be subject to homeostatic regulation. A large part of the regulation on the high side, for animals like us, is accomplished by the evaporation of water secreted through the skin. A rise in external temperature, or an increase in heat production, is followed by a slight rise in internal temperature which activates the sweating mechanism through a certain well known route. And as perspiration is evaporated, heat is removed from the body and temperature lowered. It seems pretty clear that the regulated output of the system is temperature. But while this is being regulated, what is happening to the sweating mechanism? Necessarily every bit of reduced variation in temperature is bought by increased variation in sweating. Likewise where temperature is regulated during cold, there is, on the other side of the picture, an increase in heat production. This, one may call *heterostasis*.

On an argument from the conservation of energy, the case of sweating is not a special one, but rather is typical. In general, the organism, let us say, is acted upon by a force of a certain energy. The displacement produced by that force cannot be neutralized without the use of further force, inasmuch as the energy of the new state cannot be made to vanish. In the case of homeostasis, it is the organism itself which provides the restoring energy, by increased sweating or shivering, by changes in circulation, changes in digestion, and ultimately changes in fuel consumption. In cases of changed energy relations between the organism and environment, then, homeostasis can be maintained for one variable only at the expense of heterostasis in at least one other.

The chief homeostatic systems, like that for temperature regulation, seem to have feedback circuits with rather short loops, turning back from a point near the output and opposing the system response at some point not much earlier. Yet there are some feedbacks, perhaps longer ones, which regulate the admission of energy to the organism. Such are the tympanic and iridic reflexes, and possibly certain mechanisms which affect the sensitivity of afferent nerves. By operating at the input these stabilize the whole system to some degree, and heterostasis is not so obvious. Nevertheless one must recognize that it exists. Because the system operates as an amplifier, introducing energy along the way, it is less costly to oppose system response at an early stage. Yet even here a certain amount of work must be performed: the tympanic and iridic muscles do, of course, consume energy which they would not if they were not regulating input. Their action, therefore, shows greater variation while the main auditory and visual channels are showing less.

The point seems general that an organism's constancy in one aspect implies variability in another. It is speaking rather loosely, therefore, to say that an organism as a whole maintains an internal constancy in spite of external variations. Rather, some features of its internal state are kept constant by homeostasis, some by simpler means, and some vary in compensatory fashion.

There is one final argument for the

generality of homeostasis to be weighed: the argument from its necessity for survival. If survival were absolute, the argument would be more compelling, for it would be difficult to see how anything but an *ad hoc* mechanism could counter all the vagaries of the environment. But organisms and species are both mortal, and it seems quite possible that their limited survival can occur without their having the best of imaginable mechanisms for its assurance. It is hard to say just what life span would be permitted by the expected combination of good and bad accidents.

So far the argument has concerned cases in which there is a tendency toward equilibrium somewhere. But there are physiological mechanisms which work toward the opposite end, lowering the threshold for an activity which is already under way. There is, for example, the facilitation effect described by Gellhorn and his collaborators (2): a muscular activity will lower the cortical threshold for more of the same activity. On the neuronal level it is well known that a large depolarization (spike potential) is more easily produced in a tissue which already has a small one (local bio-electric disturbance). Prior stimulation of the muscle spindles over the "gamma system" increases their susceptibility to later stimuli (4). Some of these positively accelerated phenomena depend on positive feedback, and others do not.[5] In organismic behavior there are likewise events, such as perseveration, which suggest a positive feedback. In general, a reinforcement effect, contingent upon an action and leading to its repetition, fits quite well the model of positive feedback. Learning may thus be represented by such a model. Still more generally it is quite usual that a response is far more energetic than its stimulus. Instead of resisting a sensory disturbance, the organism amplifies it. While this is not necessarily a positive feedback, it is often hard to see that it is equilibration, either. Possibly the application of the activating and positive feedback models should be explored more than it has been.

Returning to the original question, the answer is that homeostasis would not serve as a good all-around model for behavior. Using the word in a very broad functional sense will merely enlarge the vocabulary beyond necessity. If we use it in the narrow sense of a negative feedback mechanism, we find that the model does not, and indeed cannot, fit all the facts. We can make it conform by adding a series of inventions to the facts observed, but there seems no reason for doing so. Homeostasis exists, of course, with respect to certain variables, and one should try to find out what these are. But it is to be understood as a special case of the more general conception of the response of systems to inputs. There is no compulsion to think that the organism is an elaborate machine for the purpose of getting itself back to the *status quo ante*, or, indeed, for any other purpose.

REFERENCES

1. ASHBY, W. R. *An introduction to cybernetics.* New York: Wiley, 1956.
2. BOSMA, J. F., & GELLHORN, E. Electromyographic studies of muscular coordination on stimulation of the motor cortex. *J. Neurophysiol.,* 1946, **9**, 263–274.
3. CANNON, W. B. *The wisdom of the body.* (2nd ed.) New York: Norton, 1939.
4. GRANIT, R. *Receptors and sensory perception.* New Haven: Yale, 1955.
5. MILLER, J. G. Toward a general theory for the behavioral sciences. *Amer. Psychologist,* 1955, **10**, 513–531.

(Received February 4, 1957)

---

[5] A. Lundervold called my attention to the frequent occurrence of positive feedback in neural processes.

Psychological Review
Vol. 65, No. 1, 1958

HOW SHALL A THING BE CALLED?

ROGER BROWN

Massachusetts Institute of Technology

The most deliberate part of first-language teaching is the business of telling a child what each thing is called. We ordinarily speak of *the* name of a thing as if there were just one, but in fact, of course, every referent has many names. The dime in my pocket is not only a *dime*. It is also *money*, a *metal object*, a *thing*, and, moving to subordinates, it is a *1952 dime*, in fact a *particular 1952 dime* with a unique pattern of scratches, discolorations, and smooth places. When such an object is named for a very young child how is it called? It may be named *money* or *dime* but probably not *metal object*, *thing*, *1952 dime*, or *particular 1952 dime*. The dog out on the lawn is not only a *dog* but is also a *boxer*, a *quadruped*, an *animate being;* it is the *landlord's dog*, named *Prince*. How will it be identified for a child? Sometimes it will be called a *dog*, sometimes *Prince*, less often a *boxer*, and almost never a *quadruped*, or *animate being*. Listening to many adults name things for many children, I find that their choices are quite uniform and that I can anticipate them from my own inclinations. How are these choices determined and what are their consequences for the cognitive development of the child?

Adults have notions about the kind of language appropriate for use with children. Especially strong and universal is the belief that children have trouble pronouncing long names and so should always be given the shortest possible names. A word is preferable to a phrase and, among words, a monosyllable is better than a polysyllable. This predicts the preference for *dog* and *Prince* over *boxer*, *quadruped*, and *animate being*. It predicts the choice of *dime* over *metal object* and *particular 1952 dime*.

Zipf (10) has shown that the length of a word (in phonemes or syllables) is inversely related to its frequency in the printed language. Consequently the shorter names for any thing will usually also be the most frequently used names for that thing, and so it would seem that the choice of a name is usually predictable from either frequency or brevity. The monosyllables *dog* and *Prince* have much higher frequencies according to the Thorndike-Lorge list (8) than do the polysyllables *boxer*, *quadruped*, and *animate being*.

It sometimes happens, however, that the frequency-brevity principle makes the wrong prediction. The thing called a *pineapple* is also *fruit*. *Fruit* is the shorter and more frequent term, but adults will name the thing *pineapple*. Similarly they will say *apple*, *banana*, *orange*, and even *pomegranate;* all of them longer and less frequent words than the perfectly appropriate *fruit*. Brevity seems not to be the powerful determinant we had imagined. The frequency principle can survive this kind of example, but only if it is separated from counts like the Thorndike-Lorge of over-all frequency in the printed language. On the whole the word *fruit* appears more often than the word *pineapple* (and also is shorter), but we may confidently assume that, when pineapples are being named, the word *pineapple* is more frequent than the word *fruit*. This, of course, is a kind of frequency more directly relevant to our problem. Word counts of general usage are only very roughly applicable to the prediction of what will be said when something is named. What we need is referent-name counts. We don't have them, of course, but if we had them it is easy to see that they would improve our predictions. Bananas are called *banana,* apples *apple,* and oranges *orange* more often than any of them is called *fruit.* The broad frequency-brevity principle predicts that *money* and *dime* will be preferred to *metal object, 1952 dime,* and *particular 1952 dime,* but it does not predict the neglect of the common monosyllable *thing.* For this purpose we must again appeal to imagined referent-name counts, according to which dimes would surely be called *dime* or *money* more often than *thing.*

While the conscious preference for a short name can be overcome by frequency, the preference nevertheless affects the naming act. I have heard parents designate the appropriate objects *pineapple, televsion, vinegar,* and *policeman;* all these to children who cannot reproduce polysyllabic words. Presumably they use these names because that is what the referents are usually called, but the adult's sense of the absurdity of giving such words to a child is often evident. He may smile as he says it or remark, "That's too hard for you to say, isn't it?"

Some things are named in the same way by all adults for all children. This is true of the apple and the orange. Other things have several common names, each of them used by a specifiable group of adults to specifiable children. The same dog is *dog* to most of the world and *Prince* in his own home and perhaps on his own block. The same man is a *man* to most children, *policeman* to some at some times, *Mr. Jones* to the neighborhood kids, and *papa* to his own. Referent-name counts from people in general will not predict these several usages. A still more particular name count must be imagined. The name given a thing by an adult for a child is determined by the frequency with which various names have been applied to such things in the experience of the particular adult. General referent-name counts taken from many people will predict much that the individual does, but, for a close prediction, counts specific to the individual would be needed.

The frequencies to which we are now appealing have not, of course, been recorded. We are explaining imagined preferences in names by imagined frequencies of names. It is conceivable, certainly, that some of these specific word counts might be made and a future naming performance independently predicted from a past frequency. Probably, however, such frequencies will never be known, and if we choose to explain particular naming performances by past frequencies we shall usually have to infer the frequency from the performance.

BEYOND THE FREQUENCY PRINCIPLE

A frequency explanation is not very satisfying even when the appeal is to known frequencies. The question will come to mind: "Why is one name more common than another?" Why is a dog called *dog* more often than *quadruped* and, by some people, called *Prince* more often than *dog?* Perhaps it just happened that way, like driving on the right side of the road in America and on the left in England. The convention is preserved but has no justification outside itself. As things have worked out, coins are usually named by species as *dime, nickel,* or *penny* while the people we know have individual names like *John, Mary,* and *Jim.* Could it just as easily be the other way around? Might we equally well give coins proper names and introduce people as types?

The referent for the word *dime* is a large class of coins. The name is equally appropriate to all members of this class.

To name a coin *dime* is to establish its equivalence, for naming purposes, with all other coins of the same denomination. This equivalence for naming purposes corresponds to a more general equivalence for all purposes of economic exchange. In the grocery one dime is as good as another but quite different from any nickel or penny. For a child the name given an object anticipates the equivalences and differences that will need to be observed in most of his dealings with such an object. To make proper denotative use of the word *dime* he must be able to distinguish members of the referent category from everything else. When he learns that, he has solved more than a language problem. He has an essential bit of equipment for doing business. The most common names for coins could not move from the species level to the level of proper names without great alteration in our nonlinguistic culture. We should all be numismatists preparing our children to recognize a particular priceless 1910 dime.

Many things are reliably given the same name by the whole community. The spoon is seldom called anything but *spoon,* although it is also a piece of *silverware,* an *artifact,* and a *particular ill-washed restaurant spoon.* The community-wide preference for the word *spoon* corresponds to the community-wide practice of treating spoons as equivalent but different from knives and forks. There are no proper names for individual spoons because their individuality seldom signifies. It is the same way with pineapples, dimes, doors, and taxicabs. The most common name for each of these categorizes them as they need to be categorized for the community's nonlinguistic purposes. The most common name is at the level of usual utility.

People and pets have individual names as well as several kinds of generic name. The individual name is routinely coined by those who are disposed to treat the referent as unique, and is available afterwards to any others who will see the uniqueness. A man at home has his own name to go with the peculiar privileges and responsibilities binding him to wife and child. But the same man who is a one-of-a-kind *papa* to his own children is simply a *man* to children at large. He is, like the other members of this large category, someone with no time to play and little tolerance for noise. In some circumstances, this same man will be given the name of his occupation. He is a *policeman* equivalent to other policemen but different from *bus drivers* and *Good Humor men.* A policeman is someone to "behave in front of" and to go to when lost. To the kids in the neighborhood the man is *Mr. Jones,* unique in his way—a crank, bad tempered, likely to shout at you if you play out in front of his house. It is the same way with dogs as with people. He may be a unique *Prince* to his owners, who feed and house him, but he is just a *dog* to the rest of the world. A homeless dog reverts to namelessness, since there is none to single him out from his species. Dimes and nickels have much the same significance for an entire society, and their usual names are fixed at this level of significance. People and pets function uniquely for some and in various generic ways for others. They have a corresponding variety of designations, but each name is at the utility level for the group that uses it. Our naming practices for coins and people correspond to our nonlinguistic practices, and it is difficult to imagine changing the one without changing the other.

The names provided by parents for children anticipate the functional structure of the child's world.[1] This is not, of course, something parents are aware of doing. When we name a thing there does not seem to be any process of choice. Each thing has its name, just one, and that is what we give to a child. The one name is, of course, simply the usual name for us. Naming each thing in accordance with local frequencies, parents unwittingly transmit their own cognitive structures. It is a world in which *Prince* is unique among dogs and *papa* among men, *spoons* are all alike but different from *forks*. It may be a world of *bugs* (to be stepped on), of *flowers* (not to be picked), and *birds* (not to be stoned). It may be a world in which *Niggers,* like *spoons,* are all of a kind. A division of caste creates a vast categorical equivalence and a correspondingly generic name. *Mr. Jones* and *Mr. Smith* do not come out of racial anonymity until their uniqueness is appreciated.

---

[1] The equivalence of dimes and their distinctiveness as a class from nickels and pennies

Adults do not invariably provide a child with the name that is at the level of usual utility in the adult world. An effort is sometimes made to imagine the utilities of a child's life. Some parents will, at first, call every sort of coin *money*. This does not prepare a child to buy and sell, but then he may be too young for that. All coins are equivalent for the very young child in that they are objects not to be put into the mouth and not to be dropped down the register, and *money* anticipates that equivalence. A more differentiated terminology can wait upon the age of store-going. Sometimes an adult is aware of a child's need for a distinction that is not coded in the English lexicon. A new chair comes into the house and is not going to be equivalent to the shabby chairs already there. A child is permitted to sit on the old chairs but will not be permitted on the new one. A distinctive name is created from the combinational resources of the language. *The new chair* or *the good chair* is not to be assimilated to *chairs* in general.

Eventually, of course, children learn many more names for each thing than the one that is most frequent and useful. Sometimes a name is supplied in order to bring forward an immediately important property of the referent. A child who starts bouncing the coffee pot needs to be told that it is *glass*. Sometimes a name is supplied to satisfy the child's curiosity as to the place of a referent in a hierarchy of categories. Chairs are *furniture* and so are tables; carrots are a *vegetable* but apples are not. Probably, however, both children and adults make some distinction among these various names. *The* name of a thing, the one that tells what it "really" is, is the name that constitutes the referent as it needs to be constituted for most purposes. The other names represent possible recategorizations useful for one or another purpose. We are even likely to feel that these recategorizations are acts of imagination, whereas the major categorization is a kind of passive recognition of the true character of the referent.

THE CHILD'S CONCRETE VOCABULARY

It is a commonplace saying that the mind of a child is relatively "concrete"

---

is strongly suggested by the appearance of individual coins as well as by their names. Variations in size, weight, and hue are far greater between classes than within a class. This, of course, is because coins are manufactured in accordance with a categorical scheme which is also represented in our names for coins. It is possible, then, that a child might structure coins in the culturally approved manner if he never heard them named at all. However, we cannot be sure that an untutored child would not put all shiny new coins into one class and all the dingy specimens into another. When the referents are not manufactured articles but are such things as dogs, people, flowers, and insects, it is clear that autochthonous factors in perception do not force any single scheme of categorization. The names applied must be the child's principal clue to the locally functioning scheme.

and the mind of an adult "abstract." The words "concrete" and "abstract" are sometimes used in the sense of subordinate and superordinate. In this sense a relatively concrete mind would operate with subordinate categories and an abstract mind with superordinate categories. It is recorded in many studies of vocabulary acquisition (e.g., 2, 6) that children ordinarily use the words *milk* and *water* before the word *liquid;* the words *apple* and *orange* before *fruit; table* and *chair* before *furniture; mamma* and *daddy* before *parent* or *person;* etc. Very high-level superordinate terms like *article, action, quality,* and *relation,* though they are common in adult speech (8), are very seldom heard from preschool children (2). Presumably this kind of vocabulary comparison is one of the sources of the notion that the child's mind is more concrete than the mind of the adult.[2] However, the vocabulary of a child is not a very direct index of his cognitive preferences. The child's vocabulary is more immediately determined by the naming practices of adults.

The occasion for a name is ordinarily some particular thing. In the naming it is categorized. The preference among possible names seems to go to the one that is most commonly applied to the referent in question. That name will ordinarily categorize the referent so as to observe the equivalences and differences that figure in its usual utilization. There are not many purposes for which all liquids are equivalent or all fruits, furniture, or parents; and so the names of these categories are less commonly used for denotation than are the names of categories subordinate to them. It is true that words like *article, action, quality* and *relation* are rather common in adult written English, but we can be sure that these frequencies in running discourse are not equaled in naming situations. Whatever the purposes for which all articles are equivalent, or all actions or qualities, they are not among the pressing needs of children.

It is not invariably true that vocabulary builds from concrete to abstract. *Fish* is likely to be learned before *perch* and *bass; house* before *bungalow* and *mansion; car* before *Chevrolet* and *Plymouth* (6). The more concrete vocabulary waits for the child to reach an age where his purposes differentiate kinds of fish and makes of cars. There is much elaborately concrete vocabulary that is not introduced until one takes courses in biology, chemistry, and botany. No one has ever proved that vocabulary builds from the concrete to the abstract more often than it builds from the abstract to the concrete. The best generalization seems to be that each thing is first given its most common name. This name seems to categorize on the level of usual utility. That level sometimes falls on the most concrete categories in a hierarchy (proper names for significant people), and vocabulary then builds toward the more abstract categories (names for ethnic groups, personality types, social classes). Utility sometimes centers on a relatively abstract level of categorization (fish) and vocabulary then builds in both directions (perch and vertebrate). Probably utility never centers on the most abstract levels (thing, substance, etc.), and so probably there is no hierarchy within which vocabulary builds in an exclusively concrete direction.

---

[2] From the facts of vocabulary acquisition alone it is not possible to draw safe conclusions about cognitive development. Such conclusions rely on something like the following set of assumptions. A subject, whether animal or human, is ordinarily credited with a cognitive category when he extends some distinctive response to new instances of the category and withholds it from noninstances. Words, when used to denote new referents, are such a distinctive response. If children speak words they probably can make correct denotative use of them, and so the presence of the word in a child's vocabulary may be taken as evidence that he possesses the category to which the word makes reference. The instances of the category are presumed not to be differentiated by the child unless he uses words for such differentiations. If all of these assumptions are made it would seem to follow that the direction of vocabulary growth (from subordinate to superordinate or vice versa) reveals the direction of cognitive development. When the assumptions of such an argument are explicitly stated, it is clear that they are too many and too doubtful. Obviously words may be spoken but not understood; objects may be differentiated by nonlinguistic response even though they are not differentiated linguistically. However, it is not my purpose here to quarrel with these assumptions but rather to show that, even when they are accepted, the facts of vocabulary growth do not compel the conclusion that cognitive development is from the concrete to the abstract.

In the literature describing first-language acquisition (5) there is much to indicate that children easily form large abstract categories. There are, to begin with, the numerous cases in which the child overgeneralizes the use of a conventional word. The word *dog* may, at first, be applied to every kind of four-legged animal. It sometimes happens that every man who comes into the house is called *daddy*. When children invent their own words, these often have an enormous semantic range. Wilhelm Stern's (7) son Günther used *psee* for leaves, trees, and flowers. He used *bebau* for all animals. Lombroso (9) tells of a child who used *qua qua* for both duck and water and *afta* for drinking glass, the contents of a glass, and a pane of glass. Reports of this kind do not suggest that children are deficient in abstracting ability. It even looks as if they may favor large categories.

There are two extreme opinions about the direction of cognitive development. There are those who suppose that we begin by discriminating to the limits of our sensory acuity, seizing each thing in its uniqueness, noting every hair and flea of the particular dog. Cognitive development involves neglect of detail, abstracting from particulars so as to group similars into categories. By this view abstraction is a mature rather than a primitive process. The contrary opinion is that the primitive stage in cognition is one of a comparative lack of differentiation. Probably certain distinctions are inescapable; the difference between a loud noise and near silence, between a bright contour and a dark ground, etc. These inevitable discriminations divide the perceived world into a small number of very large (abstract) categories. Cognitive development is increasing differentiation. The more distinctions we make, the more categories we have and the smaller (more concrete) these are. I think the latter view is favored in psychology today. While there is good empirical and theoretical support (1, 3, 4) for the view that development is differentiation, there is embarrassment for it in the fact that much vocabulary growth is from the concrete to the abstract. This embarrassment can be eliminated.

Suppose a very young child applies the word *dog* to every four-legged creature he sees. He may have abstracted a limited set of attributes and created a large category, but his abstraction will not show up in his vocabulary. Parents will not provide him with a conventional name for his category, e.g., *quadruped*, but instead will require him to narrow his use of *dog* to its proper range. Suppose a child calls all elderly ladies *aunt*. He will not be told that the usual name for his category is *elderly ladies* but, instead, will be taught to cut back *aunt* to accord with standard usage. In short, the sequence in which words are acquired is set by adults rather than children, and may ultimately be determined by the utility of the various categorizations. This will sometimes result in a movement of vocabulary toward higher abstraction and sometimes a movement toward greater concreteness. The cognitive development of the child may nevertheless always take the direction of increasing differentiation or concreteness.

The child who spontaneously hits on the category four-legged animals will be required to give it up in favor of dogs, cats, horses, cows, and the like. When the names of numerous subordinates have been mastered, he may be given the name *quadruped* for the superordinate. This abstraction is not the same as its primitive forerunner. The schoolboy who learns the word *quadruped* has abstracted from differentiated and named subordinates. The child he was abstracted through a failure to differentiate. Abstraction after differentiation may be the mature process, and abstraction from a failure to differentiate the primitive. Needless to say, the abstractions occurring on the two levels need not be coincident, as they are in our quadruped example.

SUMMARY

Though we often think of each thing as having a name—a single name—in fact, each thing has many equally correct names. When some thing is named for a child, adults show considerable regularity in their preference for one of the many possible names. This paper is addressed to the question: "What determines the name given to a child for a thing?" The first answer is that adults prefer the shorter to the longer expression. This gives way to the frequency principle. Adults give a thing the name it is most commonly given. We have now come full circle and are left with the question, "Why is one name for a thing more common than another?"

It seems likely that things are first named so as to categorize them in a maximally useful way. For most purposes Referent A is a spoon rather than a piece of silverware, and Referent B a dime rather than a metal object. The same referent may have its most useful categorization on one level (*Prince*) for one group (the family) and on another level (*dog*) for another group (strangers). The categorization that is most useful for very young children (*money*) may change as they grow older (*dime* and *nickel*).

With some hierarchies of vocabulary the more concrete terms are learned before the abstract; probably the most abstract terms are never learned first, but it often happens that a hierarchy develops in both directions from a middle level of abstraction. Psychologists who believe that mental development is from the abstract to the concrete, from a lack of differentiation to increased differentiation, have been embarrassed by the fact that vocabulary often builds in the opposite direction. This fact need not trouble them, since the sequence in which words are acquired is not determined by the cognitive preferences of children so much as by the naming practices of adults.

REFERENCES

1. GIBSON, J. J., & GIBSON, ELEANOR J. Perceptual learning: differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.
2. INTERNATIONAL KINDERGARTEN UNION. *A study of the vocabulary of children before entering the first grade.* Baltimore: Williams & Wilkins, 1928.
3. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
4. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

5. McCarthy, Dorothea. Language development in children. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946. Pp. 477–581.

6. Smith, M. E. An investigation of the development of the sentence and the extent of vocabulary in young children. *Univer. Iowa Stud. Child Welfare,* 1926, 3, No. 5.

7. Stern, Clara, & Stern, W. *Die Kindersprache.* Leipzig: Barth, 1920.

8. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1944.

9. Werner, H. *Comparative psychology of mental development.* Chicago: Follett, 1948.

10. Zipf, G. K. *The psycho-biology of language.* Boston: Houghton Mifflin, 1935.

(Received March 16, 1957)

Bureau of Ednl. & Psy. Research Library
8-370
7.9.70

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated........ ....
Acc. No...............................

Psychological Review
Vol. 65, No. 1, 1958

"WHAT IS LEARNED?"—AN EMPIRICAL ENIGMA[1]

WILLIAM W. ROZEBOOM

University of Minnesota

Although cognitive and stimulus-response interpretations of learning have traditionally been regarded as antipodes in the theoretical disputes of behavioral psychology, it has become increasingly apparent that there are few, if any, predictive *behavioral* differences between these positions. Superficially, one might expect gross disagreement between an S-R theory, which holds to a direct attachment of responses to stimuli, and a cognitive or "expectancy" theory which maintains that conditioned stimuli elicit sensory processes, or surrogates thereof, previously associated with other stimuli. Yet when one reflects that the concept of "mediation response" (27) or "anticipatory goal response" (15) allows a stimulus to evoke a CR whose proprioceptive feedback provides a mechanism by which a stimulus may have conditioned sensory consequences, it is no longer surprising that no one has yet succeeded in cooking up a "critical" experiment to settle the controversy.

It is entirely too rash, however, to conclude with Kendler (20) that speculations as to "what is learned" are empirically meaningless. To do so is to overlook the crucial point that the known—or supposed—facts of behavior have driven S-R theorists from their older models *without* mediation responses to a more complex theoretical analysis which is, within rather wide limits, *formally* identical with expectancy theory. There *are* important predictive differences between the old, unmediated, S-R theory on the one hand, and mediated S-R or expectancy theory on the other. Hence, if by "What is learned?" we are concerned not with the connotative imagery associated with theoretical terms but with the formal properties of learning theories, it would appear that this question does have empirical significance. The fact that behavior theories of widely divergent conceptual origins appear rapidly to be converging to a common formal structure strongly implies the existence of certain *empirical* principles of behavior which necessitate this particular structure. And if such is indeed the case, we are confronted with a problem: Exactly what *are* the empirical principles which give rise to mediational theories of learning?

The question just posed is not as simple as it might appear. Every graduate student of behavioristics is familiar with, e.g., the "latent learning" experiments which were supposedly critical for expectancy theories, and whose explanations in S-R terms draw so heavily on mediation responses. But an experiment is not an *empirical principle;* by the latter, we understand a lawful relation among observation variables which, in conjunction with other empirical principles, determines the outcome of an experiment. To understand the empirical bases for mediation hypotheses, it is insufficient to know what experimental outcomes favor a mediation hypothesis. We must also know the laws which these results exemplify.

Let me elaborate on this point, for it is an important one. There appear to be two main approaches through which one can seek understanding of natural phenomena. One—the theoretic—is to hypothesize about the underlying unobserved entities or processes and their relations which are casually responsible for the gross observed effects. The other—the empiristic—is to seek discovery of a set of dependable empirical covariations, in terms of which more complex phenomena may then be analyzed. There is no need here to weigh the relative merits of these approaches.[2] What is important to note is that there is no reason to doubt that any behavioral phenomenon can be analyzed as an instance of one or more purely empirical principles, even though these, in turn, may call for explanation in theoretical terms. The importance of such "brute facts" of behavior for empiricist and theorist alike is appreciated readily enough by reflection on the position in behavioral psychology of the empirical laws of conditioning, both classical and instrumental, the facts about deprivation and activity level, etc. Such principles may be regarded as what we *know* about behavior, in contrast to what we *surmise* (though, of course, in the last analysis this is less an absolute distinction than a matter of degree), and it is this hard core of *facts*—by which I mean not merely data, which are gathered readily enough, but the much less obvious empirical covariations which these exemplify—upon which the *science* of behavioristics is erected.

---

[1] This article is based upon a dissertation submitted to the Faculty of the Division of the Biological Sciences, the University of Chicago, in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his gratitude for the generous support of the National Science Foundation, whose fellowship program has subsidized the writer's graduate studies.

The question which I have raised is whether the problem, "What is learned?", is not reflected on the empirical level as a question about the laws of behavior. I have already argued that the structural convergence of alternate learning theories does, in fact, suggest this to be so. The question still remains, however, as to whether the particular relations responsible for mediation hypotheses are merely special instances of laws already known, *or whether, perhaps, a previously unrecognized basic principle of behavior is involved.* I shall attempt to demonstrate that the latter is, indeed, the case—or more accurately, that the empirical basis for mediation hypotheses is a fundamental behavioral principle *if*, in fact, it exists. For, as will be shown, the problem at hand carries us into a new, essentially unexplored, dimension of research.

I shall not attempt to display the empirical problem of "What is learned?" through analysis of the traditional latent learning and reasoning experiments. To do so would be to become lost in a mass of obscuring details, and to suggest that the principle for which we are seeking manifests itself only in special circumstances. Rather, I shall demonstrate that *the empirical possibility which gives rise to mediation theories of learning is an intrinsic aspect of each and every instance of conditioning phenomena.* The known facts of conditioning are traditionally phrased in a way that conceals a highly critical ambiguity concerning what *are* the facts of conditioning. In the next section, I shall attempt to formulate the conditioning paradigm in complete generality, and to exhibit the correlated empirical mediation problem, without whose answer a resume of the facts of conditioning is as incomplete as a specification of geographical location which gives latitude but not longitude. But to illustrate my point and to illuminate the general analysis, I will first cite a specific example.

Perhaps the oldest empirical principle of behavioristics is that of classical conditioning. This may be expressed:

(I) *If an organism repeatedly encounters, under suitable circumstances, a stimulus,* S_c, *immediately preceding a*

---

[2] If space permitted, I would argue that theoretical concepts are—within sharply circumscribable limits—not merely justified, but are more or less *compelled* by certain kinds of empirical data.

second stimulus, S_u, *which evokes response* R, *then* S_c *will come to evoke* R.

With some qualifications (chiefly in regard to the similarity of CR and UR), (I) may stand as a close paraphrase of conventional formulations of classical conditioning, and *seems* to assert no more than is justified by generalization from known data. But compare it with the following formulation:

(II) *If an organism repeatedly encounters, under suitable circumstances, a stimulus,* S_c, *immediately preceding a second stimulus,* S_u, *then* S_c *will come to evoke the same response(s) as elicited by* S_u.

Both (I) and (II) are equally supported by the facts of conditioning, so far as these are known; *yet* (I) *implies that classical conditioning establishes a bond between* S_c *and* R *which is independent of the subsequent behavioral effects of* S_u, *whereas* (II) *implies that any modifications in the response-evocation properties of* S_u *will be passed along to* S_c*—i.e., that classical conditioning establishes a functional equivalence between conditioned and unconditioned stimulus.* Since the behavioral effects of the US have almost invariably been left unaltered in conditioning experiments, there is absolutely no justification for assuming (I) to be any more representative of the empirical facts than (II). Yet it is traditionally assumed by empiricists and S-R theorists that the conditioned reflex is a "direct" or "mechanical" attachment of response to stimulus—i.e., that (I) is the case. The truth is that we *simply don't know* what the relationship is between the responses to the CS and US subsequent to the conditioning operation, despite the fact that the value of this information, both for theoretical and applied purposes, can scarcely be overestimated.

The moral that I want to draw is that our knowledge of the empirical facts of classical conditioning is incomplete—not in the relatively trivial sense that we are not yet acquainted with all the parameters of its occurrence, but in the extremely basic sense that we have only partial knowledge of its empirical *consequences*. Moreover, unless we explicitly stipulate that the behavioral effects of the US remain constant, it appears extremely difficult to express what we do know about conditioning without inadvertently prejudging the results of experiments yet to be conducted. (Such prejudgments undoubtedly account for our traditional assumption that a conditioned reflex, once established, is independent of the subsequent effects of the US.) But the relation between the response-evocation properties of the CS and those of the US constitutes an empirical principle yet to be determined, and it is the likelihood that a significant relationship does in fact exist, along with others of its kind, that has encouraged the development of mediational theories of learning.

THE GENERALIZED CONDITIONING PARADIGM

The preceding example demonstrates how, for classical conditioning, the problem, "What is learned?", is not primarily a theoretical issue at all, but a straightforward query about the brute facts of behavior. As we shall now see, a similar problem exists for every form of conditioning. To show this in complete generality, it is first expedient to assimilate the various instances of conditioning—conditioned reflex, instrumental conditioning, conditioned reinforcement, etc.—under a single rubric.

Most, if not all, of those instances of learning to which the term "conditioning" has been applied, as well as foreseeable extensions of this term, may be subsumed under the formula of a stimulus, S_c, acquiring certain behavioral effects (for a given organism) through participation of S_c in certain contingency relations with other environmental events. By "behavioral effect" I mean any change, transient or enduring, observed or inferred, in the behavioral attributes of an organism which results from reception of the stimulus by that organism. The simplest and best known behavioral effect of a stimulus is, of course, response evocation, but more complex consequences of stimuli for behavior are also known. One such effect is (instrumental) "reinforcement," which confers upon a stimulus the ability to produce instrumental conditioning. Still another is drive arousal or reduction, which differs from simple response evocation in that the former is characterized by a change in activity level which may be expressed in a variety of behaviors. There is evidence that these effects may be acquired and lost in manners vary similar to the acquisition and loss of response evocation.

Since environmental events are presented to the organism in terms of stimuli,[3] we may formulate the conditioning schema as follows: *A phenomenon is known as "conditioning" when it may be described as the acquisition by a stimulus,* S_c, *of a behavioral effect,* E_c, *through participation of* S_c *in a contingency relation,* $T(S_c, S_u)$, *with an "unconditioned" stimulus,* S_u, *whose behavioral effects are* E_u. (Although the origins of the term "unconditioned stimulus" lie in the innate autonomic reflexes of Pavlovian conditioning, it is terminologically convenient to regard the stimulus through which conditioning is accomplished as an unconditioned stimulus, even though it may have received its own behavioral effects through a previous conditioning procedure.[4]) Let "*E* is a behavioral effect of stimulus *S*" be abbreviated "$S \rightarrow E$." Then if $S_u \rightarrow E_u$ and $T(S_c, S_u)$ together causally imply $S_c \rightarrow E_c$, $T(S_c, S_u)$ is known as a "conditioning operation," while if $S_c \rightarrow E_c$ is due to a conditioning operation, S_c and E_c are a "conditioned stimulus" and a "conditioned effect," respectively. The role of E_u in the conditioning procedure seems to be that of determining the nature of the conditioned effect, E_c, since the conditioning powers of a given S_u are altered as its E_u is altered.

It must clearly be appreciated that the generalized conditioning paradigm, so formulated, in no way asserts a *law;* it merely describes those generic empirical features which are exemplified by conditioning phenomena. So far, nothing has been said about the kinds of behavioral effects involved, nor have we specified any details of the conditioning operation, $T(S_c, S_u)$. There will be as many species of conditioning as exist behavioral effects and contingency relations with the appropriate consequences. Neither, as yet, has the relation between the unconditioned effect, E_u, and the conditioned effect, E_c, been specified; we shall turn to a detailed analysis of this following a brief summary of the two major known types of conditioning operations.

*Classical, Pavlovian, or type-*S *conditioning.* In this instance, the occurrence of the unconditioned stimulus, S_u, is made contingent only upon the occurrence of the conditioned stimulus, S_c. So far as is known, classical conditioning can be accomplished whether the behavioral effect of S_u is innate or itself conditioned, though for technical reasons, higher order classical conditioning is difficult to demonstrate empirically

---

[3] Since *behavioral* manipulations of the organism consist of *environmental* modifications, the independent variables of behavioristics, including drive operations, can always be formulated in terms of stimulus situations, although certain methodological complexities arise in the case of the deprivations.

[4] It would be preferable to call S_u the "conditioning" stimulus, but this makes for excessive confusion with "conditioned" stimulus.

(35, p. 244 f.), and is, perhaps, but weakly established (28). With the behavioral effects of S_u held constant, the conditioned effect, E_c, appears to be similar to the unconditioned effect, E_u, when E_u is response evocation, positive or negative reinforcement, or drive evocation (e.g., 7, 9, 13, 19, 24, 35). Available evidence (e.g., 6, 22) tends to weigh against the conditioned acquisition of drive reduction.

*Instrumental, or type-*R *conditioning.* When the occurrence or cessation of a stimulus, S_u, is made contingent upon emission of a response, R, in the presence of a stimulus, S_c—i.e., the occurrence or cessation of S_u is dependent on the joint occurrence of S_c and R—the subsequent likelihood of R as a response in the presence of S_c may be altered, depending upon the reinforcement value of S_u (e.g., 13, 19, 35). When the onset of S_u contingent upon R in the presence of S_c results in the increment of R as a response tendency to S_c, then S_u is known as a "positive" reinforcer. When it is the termination of S_u that strengthens $S_c \rightarrow R$, then S_u is termed a "negative" reinforcer. The onset of a negative reinforcer contingent upon R in the presence of S_c is known to decrease the strength of $S_c \rightarrow R$, perhaps through instrumental strengthening of competing responses (e.g., 11). During instrumental conditioning, the reinforcement value and probably other behavioral effects (other than drive reduction) of the reinforcing stimulus also apparently tend to become conditioned to S_c, undoubtedly through a type-S process. Thus R is only a portion of the total effects, E_c, acquired by S_c during instrumental conditioning. (Note, incidentally, that classical conditioning may be conceived as a special, or degenerate, case of instrumental conditioning in which the instrumental response, R, is the null-response—i.e., S_u is contingent upon S_c plus no particular response.) There is still much that is unknown about the parameters of instrumental conditioning, particularly in regard to the participation of drive.

THE PROBLEM OF CONDITIONED GENERALIZATION

In predicting the behavioral effects of a conditioned stimulus, it is insufficient to know merely the stimulus contingencies in which S_c has participated; we also need to know the behavioral effects of S_u. That is, E_c is a function not merely of $T(S_c, S_u)$, but also of E_u. But what do we know of this function empirically?

The extent of our knowledge here has already been summarized (perhaps overtersely) above. We know that *so long as the behavioral effects of* S_u *remain constant*, roughly $E_c = E_u$ for classical conditioning,[5] and roughly $E_c = R + E_u$ for instrumental conditioning, where R is the instrumental response (or, in certain instances, its inhibition). But this leaves unspecified *which* behavioral effects of S_u determine E_c, for, in principle, the effects of S_u can vary with time. At a given moment, are the conditioned effects of S_c determined by the effects of S_u *at that same moment*, or by the effects of S_u *at the time of conditioning?* (There are, of course, still other possibilities.) Either alternative has as much a priori likelihood as the other, and only data from experiments in which the effects of S_u are altered subsequent to conditioning will permit decision betwen them.

---

[5] Cognitive theorists have shown an inordinate concern over the apparent fact that the classically conditioned response tends to differ somewhat from the unconditioned response. There is no more a priori reason why E_c should be identical with E_u than some other fixed function of E_u at the time of conditioning—neither is any less "mechanical" than the other. The difference between E_c and E_u lends no more special credence to expectancy theories than does the even greater lack of similarity between E_c and E_u during instrumental conditioning. What *is* important for a mediation theory is whether the behavioral effect of S_c, *whatever* this may be when E_u is held constant, shows any tendency to covary with E_u after conditioning.

The analysis so far has been conducted in a quasi-formalistic manner in an effort to exhibit the problem of the functional dependence of conditioned effect upon unconditioned effect as, logically, an intrinsic aspect of conditioning phenomena. Let me now put it straightforwardly. *When a stimulus,* S_c, *is conditioned to a behavioral effect,* E_c, *through contingencies of* S_c *with an unconditioned stimulus,* S_u, *whose behavioral effect is* E_u, *what is the functional dependence, if any, of* E_c *upon* E_u *after conditioning?* For greater clarity, we may formulate the problem as an experimental paradigm.

Phase 1. Condition a behavioral effect, E_c, to a neutral stimulus, S_c, by making stimulus S_u, with behavioral effect E_u, contingent upon S_c alone, or upon a response, *R*, in the presence of S_c.

Phase 2. Extinguish E_u as an effect of S_u and condition to S_u a new effect, E'_u. Extinction of E_u is not essential, since the objective here is only to modify the effects of S_u. However, extinction of E_u, when this is possible, tends to maximize the difference in the effects of S_u before and after modification, and hence to maximize evaluation of the results in Phase 3.

Phase 3. Test S_c for its influences on behavior.

There are three major alternative results which might be anticipated as the outcome in Phase 3. These are not logically exhaustive, but would seem to cover the intuitive expectations of the behaviorist.

A. Stimulus S_c may continue to have E_c as its behavioral effect, and have gained no new effect. This indicates that under the parameters of the experiment, the behavioral effects of a CS are independent of the subsequent effects of the US. Such a result is commensurate with a nonmediational theory of learning.

B. S_c may have lost E_c as an effect, but instead, have acquired effect E'_c, where E'_c is that behavioral effect which S_c would have acquired during the initial conditioning had E'_u, rather than E_u, been the effect of S_u at that time. This would imply that, under the parameters of the experiment, the behavioral effects of a conditioned stimulus continue to be determined by those of the stimulus through which conditioning was achieved, so that, empirically speaking, the conditioning operation has established a stimulus-stimulus relationship in that the organism's responses to S_c take account, so to speak, of the current behavioral significance of S_u. Theoretically, this would imply that the relation between a CS and its effect is mediated by a process whose behavioral properties are under control of the US.

C. S_c may retain effect E_c, perhaps more or less attenuated, and also acquire some degree of E'_c, thus implying that the effects of S_c are only partly dependent upon those of S_u under these parameters. The correlated theoretical interpretation is that the behavioral effects of the CS are multidetermined, some involving a mediation process controlled by the US, and others being independent of it.

For convenient reference, it is desirable to find a distinctive title for the experimental paradigm just described. Since Outcomes B and C display a dependence of the behavioral effects of one stimulus upon those of another as the result of a conditioning procedure, the name "conditioned effect dependency" provides a literal description but is unpleasantly cumbersome. However, when the effects E_c and E'_c of the conditioned stimulus are similar, respectively, to those, E_u and E'_u, of the unconditioned stimulus, the much simpler term, "conditioned generalization," is also appropriate. For the sake of descriptive expediency, therefore, let me stretch a point and refer to the generic design as the *conditioned generalization* paradigm, even though the effects of the CS and US need not literally be the same.

There are several important observations to be made here. First, let me emphasize once again that although the conditioned generalization paradigm is distinctive as an experimental procedure, the empirical question which it is designed to answer is *not* a problem separate from or secondary to the facts of conditioning. Quite the contrary, it is a question *about* the facts of conditioning. As was illustrated earlier for the conditioned reflex, the empirical laws of conditioning are traditionally formulated in a way that negatively prejudges the existence of conditioned generalization, with the result that any experiment which appears to exemplify this effect is regarded as a special and complex instance to be explained away, if possible, in terms of more "fundamental" empirical principles. But from the available data, there is no more reason to think that the primary empirical consequence of the conditioning operation is establishment of a rigid $S_c \rightarrow E_c$ connection than it is the formation of a generalization from S_u to S_c. We simply do not yet know what the facts are.

Secondly, it is important to realize that the question of conditioned generalization is not a unitary problem to be answered by a single experiment, but a correlate to *every instance* of conditioning. There is no particular reason to think that the degree of conditioned generalization must be the same for every type of conditioning. In fact, judging from what data there are, it seems probable that the degree of conditioned generalization is not constant even for a given type of conditioning, but is highly dependent upon the particular parameters of the learning situation.

The next point is an extension of the first two. Since every type of conditioning is a conditioned generalization phenomenon (if only in a trivial sense that E_c is a constant function of E_u), the problem of conditioned generalization is quite literally a new *dimension* in the empirical study of learning, rather than merely an unexplored sector of an old framework. As such, the empirical principles which will emerge from the experimental study of this problem are bound to have powerful repercussions for our understanding of behavior. Not merely are the empirical facts of readily foreseeable applied value; data concerning the dependence of conditioned generalization upon parameters yet to be discovered will also provide a new set of observation variables intimately correlated with intra-organismic processes for which at present we have only tenuous theoretical speculations. It is perhaps not unfair to claim that conditioned generalization is one of the most important problems now confronting behavioristics.

Finally, it should be apparent that it is the facts of conditioned generalization, whatever they may be, which solve the old enigma, "What is learned?"[6] Far from being a theoretical blind alley, this is a straightforward question about the empirical consequences of the conditioning operation.

[6] Some writers (e.g., 21) have interpreted the question, "What is learned?", as concerning whether the conditioned response is a motor discharge or an achievement. However, this would seem to be more a generic problem in regard to the nature of the response than about learning as such. Since, as MacCorquodale and Meehl (21, p. 219) point out, S-R theorists are no more committed to, and in the main have no more subscribed to, a motor discharge conceptualization of the response than have cognitivists; the *dispute* over what is *learned* must lie in something other than this.

THE EVIDENTIAL BASIS FOR CONDITIONED GENERALIZATION

I have contended that it is belief in a principle of conditioned generalization that has motivated development of mediation devices such as expectancies and fractional anticipatory goal responses. Since rigorous demonstration of the dependency of theoretical mediation elements upon this kind of empirical covariation calls for a technical and fairly extensive logical analysis, I shall instead substantiate this claim by showing that a large proportion of those experiments traditionally cited as critical for mediation theories are instances of the conditioned generalization paradigm. These may be grouped into three major categories: (*a*) "secondary generalization," (*b*) "preconditioning," and (*c*) a more loosely organized set of experiments which may be called "secondary instrumental conditioning."

Secondary generalization. In 1933, Shipley (32) reported that if, for human subjects, an eyeblink is conditioned to a flash of light with a tap on the cheek as the US, and a finger flexion is subsequently conditioned to the tap (shock as the US), the light flash will be found to evoke finger flexion. On the basis of this single experiment, Hull (16, 17) proclaimed the principle of "secondary generalization," which hypothesizes that if two stimuli, S_1 and S_2, evoke the same response, R, then another response, R', conditioned to one of the stimuli, say S_1, will also be evoked by the other, S_2. This formulation differs somewhat from that of conditioned generalization in that, for secondary generalization, S_1 and S_2 have not necessarily had to participate in a common contingency relation. However, the case of secondary generalization, wherein the stimulus S_1, to which R' is subsequently conditioned, has been used previously as the US for conditioning R to S_2, as is true in the Shipley experiment,[7] is also, obviously, an instance of the conditioned generalization paradigm.

Shipley's experiment and results have been successfully repeated by Lumsdaine (13, p. 230), and a similar design by Graham (12), with dogs, also showed positive results. Thus the evidence is rather convincing that at least *some* degree of conditioned generalization is sometimes established during classical conditioning, although Razran (28, p. 328) cites certain Russian studies that cast doubt on its ubiquity. These experiments do not permit evaluation of the strength of the effect, however; in particular, they do not permit decision between alternatives B and C of the conditioned generalization paradigm, nor whether conditioned generalization is the rule, rather than the exception, for classical conditioning. It should be added that the Shipley-Lumsdaine-Graham experiments appear to be the only available evidence directly relevant to the secondary generalization hypothesis, despite the important role this supposedly empirical principle has played in mediated S-R theory.[8] The notion of secondary generalization has been occasionally invoked in other experimental contexts, such as generalization gradients among symbolically related stimuli, but more as an explanatory device than an object of experimental demonstration. Thus there are no grounds for believing in a principle of secondary generalization as distinct from conditioned generalization.

---

[7] Hull (16; 17, p. 192 f.) discusses Shipley's 1933 data as though Shipley had not conditioned finger flexion to tap on cheek, and as though transfer of finger flexion to light flash were due to shock evoking both eyeblink and finger flexion. Such an interpretation would be highly favorable to the existence of secondary generalization and the response theory of mediation, but it is definitely *not* in accord with the facts of the experiment.

[8] A later (1935) experiment by Shipley (33), in which S_1 and S_2 were separately conditioned to R, appears to support the existence of secondary generalization as distinct from conditioned generalization. However, close analysis of the control data in this and the 1933 experiment reveals the 1935 results to be ambiguous. The preconditioning experiment by Wickens and Briggs (38), mentioned later, might be interpreted as an instance of secondary generalization, but only if we are willing to regard human behavior in which a stimulus elicits a response as the result of verbal in-

Preconditioning. The preconditioning paradigm begins with repeated pairings of two supposedly neutral stimuli, S_1 and S_2, and subsequently conditions a response, R, to one of them, say S_1. Preconditioning is shown if S_2 now also tends to evoke R. The preconditioning design is a special case of the conditioned generalization paradigm in that the behavioral effect of S_1 during the pairings of S_1 and S_2 is the null-effect, or, more likely, an unidentified effect. This phenomenon was demonstrated fairly conclusively by Brogden in 1939 (2). Since then, a number of preconditioning experiments have appeared in the literature (1, 3, 4, 8, 18, 34, 38), all but one (3) reporting positive findings. But, the claims of Bitterman (1) notwithstanding, any attempt to evaluate the magnitude of the effect strongly suggests that the phenomenon is a minor one, at least judging from the published studies. Thus the empirical importance of preconditioning is still problematic.

Secondary instrumental conditioning. In this category, we may include those experiments which, following establishment of an instrumental response, modify the behavioral effects of the reinforcing stimulus. More specifically, the problem may be phrased: If an instrumental response is conditioned to a stimulus, S_c, through reinforcement by a stimulus, S_u, and the reinforcement value of S_u is now altered, will the response to S_c remain unaffected, or will it have adapted in a manner commensurate with the new behavioral significance of S_u? When S_u is positively reinforcing during conditioning, and is subsequently extinguished of positive value or aversively reconditioned, this design may also be called "secondary extinction" (29), "latent extinction" (31), or "nonresponse-extinction" (10). A typical secondary extinction experiment conditions, say, a position habit in a T maze, one goal box of which contains food or water. The animal is then placed for a while directly in the positive goal box, which is now empty of the primary reinforcer, and where the animal may also be shocked. The experimental question is whether the position habit will appear weakened on the next trial. The habit may be analyzed as a response chain held together by secondary reinforcement deriving from the sequence of discriminative cues. Reconditioning directly in the goal box tends to extinguish or aversively recondition the terminal reinforcers of the sequence, and the next maze trial reveals the immediate effect on earlier links of the response chain.

An experiment of this sort was first proposed by Tolman in 1933 (37), who obtained negative results. But shortly thereafter, Miller (23) reported that shocking rats placed directly in a distinctive goal box inhibited a running response which had previously terminated in that goal box. Secondary extinction research subsequently languished until the appearance of a number of recent studies (5, 10, 26, 29, 30, 31), most of which, unfortunately, have technical flaws (cf. 26, 29) which vitiate their interpretive significance. However, the evidence seems fairly definite that, under at least some parameters of conditioning, a lever-pressing habit in rats is quite independent of the subsequent

---

structions as an instance of "response evocation" in the same sense that we apply this phrase to infra-human behavior.

value of the reinforcers through which the habit was established (29).

An important variety of traditional latent learning research (*type* 5 [21, p. 211]) may also be reduced to the secondary-instrumental-conditioning rubric. In this instance, the organism is given experience with a maze, with eliminable reinforcers either absent, or equally distributed to all end boxes. The animal is now introduced directly into one end box and given a previously unencountered reinforcement, either positive or negative. The problem is whether or not the animal's maze behavior will have adapted to the changed conditions on the next trial. This differs from more obvious cases of secondary instrumental conditioning only in that the role in establishment of initial maze habits played by the stimuli whose secondary reinforcement values are subsequently modified is not so clear. When no known reinforcers are obviously influential in the initial maze behavior, this design may be described as "instrumental preconditioning," since it involves a stimulus, S_u (say, an end box), as contingent on a response, R, in the presence of another stimulus, S_c (the antecedent portion of the maze), but where S_u has no readily identifiable reinforcement value prior to the reconditioning operation. A number of experiments of this sort have been reported (see 21, p. 211, for summary; also, 14, 25, 36), with outcomes about equally divided between positive and negative. Thus again, the evidence appears to favor the existence of conditioned generalization under some circumstances, but more explicit evaluation of the empirical significance of these experiments is precluded by the difficulties in giving an unambiguous account of the initial maze learning in terms of conditioning principles.

In summary of the available evidence concerning conditioned generalization, it would thus appear that while the quantity of relevant data is very meager, they distinctly imply that parameters of conditioning exist under which at least *some* degree of correlation between the behavioral effects of CS and US is established. Unfortunately, the existent experiments happen, for the most part, to be of special designs which either employ a minimal or unknown unconditioned effect, E_u, or do not compare the relative strengths of E_c and E'_c subsequent to reconditioning of S_u. It is only when E_u and E'_u are strong and well defined, and the alternatives E_c and E'_c are distinct and quantifiable, that we can decide whether the *primary* result of a conditioning operation is establishment of a full-fledged generalization from US to CS, or whether the prevalent notion that conditioning attaches a specific effect to the CS is basically correct, and apparent cases of conditioned generalization are secondary phenomena to be explained as special instances of more fundamental empirical principles. It is earnestly to be hoped that the present analysis, by stripping from the empirical question of "What is learned?" its obscuring theoretical superstructure, and developing a general experimental paradigm which isolates the essential features of the problem, will stimulate research intended not so much to test a theoretical prediction as simply to learn the facts.

SUMMARY

I have contended that, far from being a blind alley, the question "What is learned?" is an extremely important empirical problem about the consequences of conditioning, and one which has been concealed by improper formulations of the known facts. We have traditionally assumed that a conditioning operation attaches a fixed behavioral effect, E_c, to the conditioned stimulus, S_c, whereas in fact, the conclusion that S_c will give rise to E_c subsequent to conditioning *is empirically justified only so long as the behavioral effects of the unconditioned stimulus,* S_u, *remain constant.* The question then arises as to the extent to which, subsequent to the conditioning operation, the effects of S_c are a function of those of S_u. I have elected to identify this as the problem of "conditioned generalization," since, in one of its forms, it asks whether or not a conditioning operation establishes a generalization from US to CS.

Since the empirical consequences of the conditioning operation constitute a problem extending over the entire range of conditioning phenomena, and one which has remained unexplored in the overwhelming majority of conditioning experiments, conditioned generalization defines a new dimension of behavioral research. What few data are available suggest that conditioned generalization does exist as a phenomenon, but are totally inadequate as a basis for judging whether it is a basic behavioral principle or merely a secondary effect. Because of the practical value of this information, and its crucial significance for the development of behavior theory, it is to be hoped that the conditioned generalization problem will soon be subjected to a concerted experimental attack.

REFERENCES

1. Bitterman, M. E., Reed, P. C., & Kubala, A. L. The strength of sensory preconditioning. *J. exp. Psychol.,* 1953, **46**, 178–182.
2. Brogden, W. J. Sensory preconditioning. *J. exp. Psychol.,* 1939, **25**, 323–332.
3. Brogden, W. J. Tests of sensory preconditioning with human subjects. *J. exp. Psychol.,* 1942, **31**, 505–517.
4. Brogden, W. J. Sensory preconditioning of human subjects. *J. exp. Psychol.,* 1947, **37**, 527–539.
5. Bugelski, B. R., Coyer, R. A., & Rogers, W. A. A criticism of pre-acquisition and pre-extinction of expectancies. *J. exp. Psychol.,* 1952, **44**, 27–30.
6. Calvin, J. S., Bicknell, E. A., & Sperling, D. Effect of a secondary reinforcer on consummatory behavior. *J. comp. physiol. Psychol.,* 1953, **46**, 173–175.
7. Calvin, J. S., Bicknell, E. A., & Sperling, D. Establishment of a conditioned drive based on the hunger drive. *J. comp. physiol. Psychol.,* 1953, **46**, 176–179.
8. Chernikoff, R., & Brogden, W. J. The effect of instructions upon sensory preconditioning. *J. exp. Psychol.,* 1949, **39**, 200–207.
9. Danziger, K. The operation of an acquired drive in satiated rats. *Quart. J. exp. Psychol.,* 1951, **3**, 119–132.
10. Deese, J. The extinction of a response without performance of the choice response. *J. comp. physiol. Psychol.,* 1951, **44**, 362–366.
11. Dinsmoor, J. A. Punishment: I. The avoidance hypothesis. *Psychol. Rev.,* 1954, **61**, 34–46.
12. Graham, D. T. Experimental transfer of conditioning in dogs. *J. exp. Psychol.,* 1944, **34**, 486–493.
13. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.
14. Honzik, C. H., & Tolman, E. C. The perception of spatial relations by the rat: a type of response not easily explained by conditioning. *J. comp. Psychol.,* 1936, **22**, 287–318.
15. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.,* 1931, **38**, 487–506.
16. Hull, C. L. The concept of the habit-family hierarchy and maze learning. *Psychol. Rev.,* 1934, **41**, 33–54, 134–152.
17. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
18. Karn, H. W. Sensory preconditioning and incidental learning in human subjects. *J. exp. Psychol.,* 1947, **37**, 540–544.
19. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
20. Kendler, H. H. "What is learned?"—a theoretical blind alley. *Psychol. Rev.,* 1952, **59**, 269–277.
21. MacCorquodale, K., & Meehl, P. E. "Edward C. Tolman." In W. K. Estes *et al., Modern learning theory.* New York: Appleton-Century-Crofts, 1954.
22. Miles, R. C., & Wickens, D. D. Effect of a secondary reinforcer on the pri-

mary hunger drive. *J. comp. physiol. Psychol.*, 1953, **46**, 77–79.

23. Miller, N. E. A reply to "Sign-Gestalt or conditioned reflex?" *Psychol. Rev.*, 1935, **42**, 280–292.

24. Miller, N. E. "Learnable drives and rewards." In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

25. Minturn, L. A test for sign-Gestalt expectancies under conditions of negative motivation. *J. exp. Psychol.*, 1954, **48**, 98–100.

26. Moltz, H. Latent extinction and the reduction of secondary reward value. *J. exp. Psychol.*, 1955, **49**, 395–400.

27. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

28. Razran, G. A note on second-order conditioning—and secondary reinforcement. *Psychol. Rev.*, 1955, **62**, 327–332.

29. Rozeboom, W. W. Secondary extinction of lever pressing behavior in the albino rat. *J. exp. Psychol.*, 1957, **54**, 280–287.

30. Scharlock, D. P. The effects of a pre-extinction procedure on the extinction of a place and response performance in a T-maze. *J. exp. Psychol.*, 1954, **48**, 31–36.

31. Seward, J. P., & Levy, N. Sign learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39**, 660–668.

32. Shipley, W. C. An apparent transfer of conditioning. *J. gen. Psychol.*, 1933, **8**, 382–391.

33. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 571–591.

34. Silver, C. A., & Meyer, D. R. Temporal factors in sensory preconditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 57–59.

35. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

36. Strain, E. R. Establishment of an avoidence gradient under latent-learning conditions. *J. exp. Psychol.*, 1953, **46**, 391–399.

37. Tolman, E. C. Sign-Gestalt or conditioned reflex? *Psychol. Rev.*, 1933, **40**, 246–255.

38. Wickens, D. D., & Briggs, G. E. Mediated stimulus generalization as a factor in sensory preconditioning. *J. exp. Psychol.*, 1951, **42**, 197–200.

(Received April 7, 1957)

Psychological Review
Vol. 65, No. 1, 1958

PERCEPTION: EVOLUTION OF A CONCEPT[1]

WILLIAM BEVAN

Emory University

This paper is essentially an examination of theories of perception—especially in their relation to general behavior theory. If it at times appears to be little more than verbal masonry, this is because defining such a relationship is, in many respects, at present a semantic problem resolvable in terms of personal preference.

Reference to general behavior theory typically implies formulations like Hull's behavior principles applied to maze learning, the learning of nonsense syllables, brightness and form discrimination, social learning, emotional disturbance, and certain perceptual phenomena, or the Gestalt principles of perceptual organization applied as an analogue to explain problem solving, and even the organization of overt responses. These contrast sharply with theories like Granit's dominator-modulator theory, and the Wever-Bray volley-principle which were devised to account for sense-organ function.

The relation of perceptual to general behavior theory. The nexus between the two classes of theory under consideration is typically delineated on a background of metaphysical bias, though this is often confused by semantic issues. Some theorists have contended, as did the early behaviorists, that behavior is objective (publicly observable) and physical (involves muscles and glands), while perception is subjective (accessible only to the pereceiver) and mental (involves conscious experience); and, since scientific theories require objective data, perceptual theories *ipso facto* are not scientific. Thus there is no inter-theory relationship to describe. This position is predicated on the assumed synonymity of "objective" and "publicly observable." An alternative position is to include as objective inferred events that stand in a logically consistent relationship with the publicly observable. This means that the objectivity of a concept is determined by the efficiency with which it yields *reliable* replications of observations (9). Perception thus qualifies as an aspect of behavior, though always an inferred aspect. Indeed, if one extends this line of reasoning, one must conclude that no instance of behavior is directly observed. Rather, it is always inferred by an intelligent observer. It is meaningful (i.e., evident as behavior) to the extent that the data of observation are similar in some fashion to data which the observer has previously used in constructing his conception of the world (1). It is, of course, not incompatible with this view to argue that gross bodily movements, since they do not require transformation by the subject into verbal form before becoming available data, are more objective than perceptual data.

Having committed oneself to this latter view—as I have—one is still faced with the necessity of defining the relationship between perception and other aspects of behavior. The early psychophysicists viewed perception as a simple mediating process—a constant in the transformation of information received by the organism from a form that exists external to the organism to a form that is utilizable in overt response. Thus considered, there is no problem of perception beyond that of translating into psychological terms—or, as some prefer, into physiological terms—the characteristics of the external world. Once the constants of transformation have been determined (cps expressed as mels, db as sones, millilamberts as brils, grams as vegs, texture gradient as slant, etc.), the perceptionist's task is complete. Indeed, since perception does not dynamically affect other aspects of behavior, one may ignore perception and correlate stimulus input with some aspect of overt response.

---

[1] Parts of the present paper were presented to the Graduate Psychology Colloquium of Purdue University and as part of a symposium on behavior theory at the second annual meeting of the Southeastern Psychological Association in the spring of 1956. I am indebted to Harry Helson and W. F. Dukes for helpful critical evaluations of the manuscript.

Recent observers insist, however, that perception is a function not only of present stimulus input and receptor function, but also of motivation (which has stimulus properties) and past stimulation—what Helson (34) calls *residual stimuli*. Thus the perceptual mechanism is a complex interactive one, and the identification and definition of the functions that influence it become indispensable to an understanding and prediction of both perception and overt response.

That understanding perception is important to an understanding of overt behavior is an insight shared by modern phenomenological and behavior theorists. Approximately 25% of the articles in the PSYCHOLOGICAL REVIEW over the past five years have been on perception. Floyd Allport (4), known for years as an uncompromising behaviorist of the unsentimental type, has written in the preface of his volume reviewing perceptual theories: "In them, if they are taken all together, lie the essentials of our present knowledge of one of the most important activities of living beings." In similar vein are Woodworth's assertion that "the conditioning experiment is really concerned with the establishment of a new perception" (73, p. 124), and Hilgard's statement that "the learning theorist must be aware of the learning within perception as well as the role of perception in learning" (39, p. 332). One may even suspect that Skinner (60) has made a concession of sorts when he writes of the "interpreted stimulus."

Aim of the present paper. Having assumed a metaphysical bias concerning the status of perception in psychology, I might follow any of several paths of discourse for the remainder of this paper. I have simply chosen to review critically some ideas about the nature of perception that have been prominent during the last fifty years, at the same time recognizing that thumbnail sketches of theoretical orientations involve grave risk of distortion and oversimplification.

AN OVERVIEW OF CONCEPTUAL DEVELOPMENT

I have titled these remarks "Perception: Evolution of a Concept." It may be countered that while change in conceptualization is evident, this does not constitute evolution, and that while there is increased complexity in the concept's definition, this is not the unidirectional unfolding which is characteristic of evolution—it would even appear that some of our notions have cycled full circle. Indeed, close examination of the recent history of thinking about perception prompts the suggestion that these changes might more properly be described as spiraling than as circling. For example, early definitions were empiricist only in the broad common-sense fashion that the metaphysics of Locke *et al.* was empiricist. In the early behaviorist's attempts to apply the systematization of science to psychology, the problem of perception was solved by eliminating the concept. Later systemizations (e.g., Hebb's) have been increasingly rigorous, and at the same time have dealt with phenomena covered by the concept. Indeed, in Hebb's system, perceptual development has central status.

When physiologizing in psychology was in vogue 35 years ago, perception was rejected as a mental phenomenon. Today, physiologizing is again in vogue, but physiological systems attempt to accommodate the behavioral facts of perception. Finally, early in the period we are considering, the role of experience in perception was broadly defined and accepted as a self-evident metaphysical principle. Later, it was repudiated by the demonstrations of the Gestalt movement. More recently, the importance of past experience has been reaffirmed and the nature of the relationship more precisely defined, this time through methodical experimentation.

Conceptual evolution can be seen, I think, in numerous ways. First, there has been an increasing sophistication of perceptual theory qua theory. Fifty years ago, perception was essentially defined by a single a priori proposition. Today, in any of the several theoretical orientations extant there are attempts to represent it by a logically related set of propositions, connected through explicitly specified operations to the level of empirical generalization. Inspect, for example, the later hypothesis theory of Bruner (13), and Postman (56). Admittedly, perceptual theory has not achieved the status of a physical theory like the dynamical theory of gases. Yet the recent successes of adaptation level in quantitative prediction make for extreme optimism (34).

Secondly, the purpose of perceptual investigation has come to include greater concern for the elaboration of theory through empirical study. Earlier, perception was a term used in the description of the content of consciousness. It represented the compounding of sensations and images of past experience. Later the study of perception meant a catalogue of demonstrations and principles designed to explain, as Koffka put it in his *Principles* (46), why things appear as they do. Still more recently, the aim has been to construct a class of empirically based propositions that may be fruitfully used—usually along with other classes of proposition—in the prediction of a wide repertoire of behavioral activities.

Thirdly, there perhaps has been a more clearly conscious appreciation of the role and status of models and a more discriminating use of them. One need contrast only the psychological atomism of Titchener (62) with the current physiological atomism of Hebb (32) or the application of physical field theory by the Gestaltists in the 1920's with its present use in Köhler's work on visual figure currents (49, 50).

Perceptual models have, for the most part, been material models. Even in psychophysics and in adaptation-level theory where the explicit model is formal, one can find, I believe, an implicit material model. Early in our fifty-year period, models were essentially static-descriptive. In modern psychophysics (31), in Gibson's psychophysics of form perception (26), and in adaptation-level theory (34) the models are of these types. (It is not intended to imply that these are of a lower order than dynamic models, for, as Helson has pointed out, the trick in science is to express dynamisms mathematically and therefore statically.) But as perceptual theories became more functional—i.e., as interest is turned toward the relation between perception and other aspects of behavior—models became more frequently dynamic models: e.g., those of Woodworth (73), Tolman (64), Brunswik (17, 19), Ames and Cantril (e.g., 21), Bruner and Postman (e.g., 13), and Vernon (65). Finally, the early models were molecular. As the holistic properties of perception were emphasized, and later the status of perception as behavior stressed, the use of molar models become prevalent (contrast the psychological atomism of Titchener [62] and the functionalistic theory of Brunswik [17]). As interest returns toward physiologizing and toward dimensional analysis in some areas of investigation and theory construction, molecular models are again coming into use—e.g., the cell-assembly theory of Hebb (32).

Fourth, the concept of perception has changed in complexity, and definition has shifted from essence to process. Early conceptualization centered around simple sensory qualities and the memory images which are vestiges of them. These elements were related in an uncomplicated fashion to a rather restricted number of physical stimulus dimensions. In more recent formulations, a larger number of variables are included, and the nature of the intervariable relationships is regarded as more complex. The relational character of perception is stressed, intermodality effects and the role of background described, and the contribution of nonsituational—i.e., motivational and learning—variables is emphasized. The view has changed from one of perception as a static picture of the external world to perception as the resultant of a complex repertoire of interacting processes.

Finally, one gains the impression of an increased variety of media used in current perceptual studies. While the majority of investigations still employ adult human subjects, animals and children appear to be used more frequently.

SOME PERCEPTUAL THEORIES IN OUTLINE

The following sketches are arranged into four broad categories, roughly chronological. This, however, is not to be viewed as a rigorous classification, for many of these views overlap in some fashion or other. The reader should expect a discussion of the place of perception in recent learning theory. However, we shall not deal formally with this matter, for learning theorists have done little more than recognize, albeit to different extents, the importance of taking cognizance of perceptual variables. Skinner (60) dismisses perception as conditional seeing, by drawing a distinction between discriminative behavior and response to discriminative behavior. Berlyne (8) proposes a place for perception in Hullian theory as an intervening stimulus-producing response (attention is its momentary effective reaction potential), and Wickens (71) regards it as dependent upon the differential reinforcement of certain magnitudes within the heterogeneous stimulus complex confronting the organism. For Tolman (64) perception is one of three types of intervening-variable systems, characterized by categorizing dimensions and intensitive dimensions (qualitative and quantitative orders). Thus, unlike others, he provides a formal distinction between perception and other mediating processes, but unfortunately he does not make clear how these formal distinctions relate to the level of empirical observation. Meanwhile, no theorist has yet reconciled the many facets of perceptual data within his model for learning.

I. Perception as Conscious Content

Titchener's views. Titchener, in his 1910 textbook (62), described two types of perception—pure and mixed. Pure perceptions are multidimensional sensations, combinations of simple sensations under the laws of attention. While the end result might appear to be unique and unanalyzable (e.g., the flavor of a peach), it is nothing more than the sum total of sensations evoked by various aspects of the stimulus object. Pure perceptions are rare, and could occur only in a truly naive perceiver, for they involve the grouping of sensations under the laws of attention. Mixed perceptions are more common and also more complex, for they involve the addition of images of past experience. These latter originally may be kinesthetic images, but as the organism matures and his experience increases they derive from all modalities. Furthermore, perceptions differ from sensations in more than complexity; they have meaning. Meaning in turn derives from context, that is, from the setting in which sensations occur. It is thus clear that meaning depends on the contiguous occurrence of more than one sensation. Meaning is an inherent property of perception, for the context that accrues from a situation is always the psychological equivalent of its logical meaning. Meaning is carried by images as well as sensations and, indeed, these latter may degenerate into a kind of symbolic shorthand, largely unconscious.

In this view the primary concern is with the content of awareness. While perception involves what the Gibsons (29) have called enrichment, its literal aspect (the perception of the material properties of things) depends on sensory processes, and thus is an invariant and faithful representation of that part of the external world with which it is concerned. The aim of psychological investigation is the analysis of experience through introspection. This, in any final analysis, is an exercise in categorizing, for the method provides no means of empirical verification or check of internal consistency. The goal of psychology is met when these analytical descriptions are complete. The question of etiology and development as an empirical problem never really arises.

James' views. William James' (45) conception of perception, like that of Titchener, is an enrichment theory. Like that of Titchener, it is concerned with the content of awareness. Perception is the consciousness of particular materially present objects. Although sensations cannot be identified except in a logical sense, perceptions are represented as the combination of sensations and reproduced attributes. Sensations have both intensity and extensity, "thingness" depending on the coalescence of different sensations. But what we find here that we do not find in Titchener is emphasis upon the constructive role of past experience. Every perception is acquired, and, indeed, as it is found in the adult, represents a long course of development of order out of chaos. Furthermore, every perception is of definite and probable things. We perceive the thing—and this point is supported by extensive reference to illusions—"which would most usually have given us that sensation" (45, p. 316). In fact, habits of perceiving become so thoroughly established that new and alternative perceptions are more and more difficult to achieve. "Old-fogeyism," James contends, "is the inevitable terminus to which life sweeps us on" (45, p. 328).

There is much in James that we see highlighted in later interpretations. The long course of development represented in perception has been emphasized by Hebb (32). James' notions of probable perceptions and old-fogeyism are brilliant anticipations of Brunswik's probability learning (15, 18), Ames and Cantril's assumptions (e.g., 21), and Duncker's functional fixedness (24). But he does not take the further step that the modern functionalists take by regarding perception as resulting from an interaction between perceiver and perceived. Perceptions are learned, but it is not clear that this involves anything more than the passive compounding of sensations and memories of sensations, providing a kind of catalogue of pictures, each available for view on appropriate occasion. This half-way view is perhaps an inevitable by-product of James' concern with content. Meanwhile, his discussion of illusions constitutes an armchair preview of the type of demonstration that Ames and his followers have popularized in the last decade.

The Gestalt orientation. The Berlin Gestaltists were among the earliest to recognize the importance of asking why objects appear as they do. Their answer was simple and direct. Taking their lead from Ernst Mach's contention that what we perceive are things (48), they reacted against the elementalism of the structuralist, arguing that sensations are logical abstractions and not psychological phenomena. Thus, if perceptions are not constructed from sensations, they must be immediately given in experience, our nervous system organizing them out of what Einstein has called the "rabble of the senses." There is assumed to be an isomorphism between brain events and experience, the structure of perception mirroring the structure of cortical organization. The Gestaltists' rejecting of the elementalism of the structuralists also means a rejection of their empiricism. Koffka (46) completes the discussion of almost every major problem of perception (e.g., figure and ground, phenomenal constancy) with a dialectical refutation of empiricism. This, however, does not mean denying the possible significance of past experience. Both Wertheimer (69) and Köhler (47) recognize it as a variable, but object to its indiscriminate application as a universal explanatory principle. While the phenomenological description of perceptions as wholes suggests a preoccupation with conscious content, the Gestalt theory is basically neurophysiological in orientation. While meaning is a complex phenomenon, depending, as Köhler has noted (48, p. 73), on personal biography, certain meaning (that of object qua object) is given in the perception, and thus by the brain dynamics. Indeed, the problem of meaning is probably what took the Gestalists away from peripherally biased structuralism to their theory of cortical organization. Curiously, their theory is still one of the enrichment type. The important transition in thought is that perceptions, as the psychological counterpart of brain dynamics, are constructs, or, more correctly, operational links between physiological constructs and the level of empirical data. Gestalt demonstrations are the means by which the functional properties of central mechanisms are known. This is most clearly illustrated in Köhler and Wallach's monograph (49) on figural aftereffects.

II. Perception and Modern Psychophysics

Graham's behavioristic psychophysics. Modern psychophysicists are, by and large, S-R psychologists. Beyond the assertion that their basic data are simple discriminations, they leave the metaphysical issues to someone else. With a few brilliant exceptions (e.g., Helson's work on adaptation level), their aim is an encyclopedia of quantitative relations between certain aspects of stimuli, defined as energy functions, and responses, similarly defined. As Graham points out (31), psychophysical functions are special cases of the general positivistic formula, $R = F \ (a \ldots z)$, where all variables are constant except stimulus dimension a. Special pains are taken to create situations which yield quantitative, easily interpreted results. There is no expectation that universal concepts can be generated.

Several general criticisms have been leveled at the psychophysicists. First, it has been contended that the tradition of sensory psychology from which they spring has blinded them to the importance of central variables in behavior in the setting of the psychophysics experiment. Second, this same background tends to result in an overly simple notion of the basic determinants of perception. Third, the restricted designs demanded by psychophysical methodology preclude identification of interactions among the various determinants of perception, and thus limit the possibility of fruitful theorizing that might stem from the data that complex designs yield. Lastly, conditions of observation are so artificialized in order to meet the aims of precise control that the data yielded have only limited applicability.

Such criticisms are basically statements of an opposing bias. Whether they are more than this waits on an empirical check of their validity. The psychophysicist must also, of course, answer them at the level of concrete data. Meanwhile, the kind of definition of basic elements that this sort of approach gives is a necessary first step in the construction of a general theory.

Helson's adaptation-level theory. Examination of Helson's work over the last 25 years shows it to be shaped by interests in the Gestalt definition of the problems of perception and in the psychophysicist's aim of quantification. His recently introduced concept of adaptation level reflects particularly this latter influence. This he presents as the basis for a general theory of perception (34, 35). In outline, it may be regarded as the basis for a general theory of behavior.

Adaptation-level theory begins with the assumption that the perceptual response may be represented as a function of the difference between some present stimulus process and an internal standard derived from a pooling of the effects of previous stimulation. The status of this standard, expressed in terms of some quantifiable property of the external stimulus situation, depends both upon the characteristics of stimuli (e.g., frequency of presentation, magnitude, nearness in time, etc.) and of their background. It is given expression as a mathematical equation. Thus the model employed, unlike those employed in the majority of current perceptual theories, is formal rather than material, although it may be suggested that behind the formal model is the notion of the organism as a logarithmic computer. Since the model is mathematical, the work which it has inspired is necessarily static-descriptive in nature. But the concept is functional and thus evolutionary and dynamic in intent.

Adaptation-level theory has functioned as a precise predictive tool in a variety of situations to which it has been applied in the last 10 years. It has, for example, provided for the quantification of the shift in the neutral point with a change in standard in the classical psychophysical weight-lifting situation (33, 34). It has successfully reconciled phenomena like color constancy, contrast, and conversion. It has made possible the quantification of concomitant changes in hue, brightness, and saturation (38). It provides a rational basis for the understanding of anchor effects, and may be applied fruitfully to the study of figural aftereffects (44). Recently, Helson has used it to scale the value of verbal terms with quantitative denotations, to demonstrate the effect of the context in which the terms are used upon these values (37), and to investigate the effects of personal and group factors upon the expression of attitudes (36).

Helson regards adaptation level to be the means for translating the notion of perceptual organization into objective and quantitative terms. In principle it does this. Successful application to the more complex situations that characterize everyday life is thus basically an empirical matter of identifying the variables and their properties that are important for organization—that is, of delineating in greater detail the three broad categories of factors that he identifies as stimulus, background, and residual. We are at present particularly short of knowledge of the nonstimulus variables. We have yet, for example, adequately to establish the criteria for differentiating relevant from nonrelevant stimulus input. When this is accomplished, A.L. provides a means of predicting, within the class of relevant variables, what constitutes equivalent stimuli. Before the advent of A.L. theory this could be accomplished only a posteriori, in empirical terms (10). Certain of the parameters important to A.L. are not yet unequivocally defined. It is not always sufficiently clear, for instance, what is background and what is residual stimulation.

Adaptation level is one of several concepts currently directed toward the aims of general theory. Its assumption of pooling is also found in Brunswik's probability learning (17). If it is redefined as a probability function, it is seen to resemble strongly Postman and Bruner's concept of hypothesis (13), and it is a quantitative version of Cantril's assumptive world (21). It is hoped that the theory will be expanded so that it will consist not only of the A.L. formula, but also of an entire set of related mathematical expressions that will give more precise formal definition not only to the reference level, but to the range—with reference to both intensity and time—of relevant stimulus input, the complexity of relevant variables, the rigidity-plasticity of the judgmental system, the stability of the reference level, the fineness of articulation within the judgmental dimensions, and other parameters that constitute the perceptual frame of reference. Just how many equations would constitute the formal set would depend, of course, on the nature of the phenomena to which the theory was applied. Such a set of equations would be analogous in function to the set that relates work to force and distance, effort to distance and resistance, mechanical advantage to effort and resistance, and power to work and time in Newtonian mechanics, or to the set that relates pressure, volume, and temperature in the dynamical theory of gases.

Gibson's psychophysics of surface and edge. Gibson is concerned primarily with literal perception (27), that is, the perception of objects qua objects. Furthermore, his objects are the solid objects of the three-dimensional visual world (26). Awareness of these, he assumes, is immediately given by virtue of a kind of isomorphism between the spatial properties of the visual stimuli and the retinal events to which they give rise. Two-dimensionality is thus a learned refinement of three-dimensional perception. One sees in Gibson's theory, as in Helson's, evidence of two major influences: the quantitative aim of psychophysics and the relational approach of Gestalt psychology. He envisages quantitative description where the elements are surface and edge rather than points of color. For him stimulation is ordinal, that is, defined by gradient relationships within the retinal process.

There is an explicit nativism in Gibson's approach, for perception is regarded as reflecting directly the properties of the retinal correlate. But, unlike the Berlin Gestaltists, Gibson stresses the correspondence rather than the noncorrespondence between the external environment and our perception of it. Learning for him is specificity learning: the gradual awareness of greater detail, the enhanced capacity for discrimination (29). Through learning, perceptions acquire meaning. But perceptions must be structured before meaning can be acquired, although some meanings may be embryonic in the same way that literal perception is. This enrichment of perception by meaning, however, is not, in the strict sense, perceptual learning.

It is curious that while Gibson's insistence on a psychophysics of object perception spells objective science, his distinction between the visual field and the visual world is pure phenomenology. His most valuable contributions are the notion of ordinal stimulation and his geometry of continuous surfaces. He has indeed made it clear that the retinal correlate is *not* a picture of the external world. Meanwhile, it may be contended that he has overestimated the veridicality of perception by his assumption of an identity between perception and judgment that is open to question (cf. his studies on size constancy [26]). Certainly his insistence that perceptual learning is specificity learning requires a distinction between perception and conception that cannot confidently be achieved beyond the verbal level. Finally, his theory is a logical construction that still awaits systematic experimental evaluation.

Attneave's application of the information theory model to form perception. The persistent need to quantify perceptions typical of everyday life has, in the last several years, served to motivate attempts to apply the information-theory model to these phenomena. Attneave has been a leader in this approach (5, 6). He suggests (5) that if perception is thought of as information handling, it can be immediately seen that much of the information provided is redundant. If the visual field were fractionated by a grid, redundancy would be found in connection with areas of homogeneous color, persistent direction or slope of contour, symmetry of design, or uniform gradient of texture—indeed, any homogeneity of pattern. The Gestalt notion of good figure thus refers to some high degree of internal redundancy. The perceptual mechanism operates to minimize redundancy, encoding input into more economical form. Hence, identifiable forms may be communicated by simply presenting contour points of maximal curvature.

This view, like Gibson's, assumes an isomorphism between the spatial properties of the external object and the organism's perception, the faithfulness with which the perception matches the object depending upon the efficiency with which the perceptual mechanism scans and encodes the input. The theory thus implies an intact, completely developed scanner, the character of the perception reflecting essentially the character of the input. As it currently stands, the model deals with redundancy associated with the geometrical properties of the speciously present stimulus. There is yet no extension to the problems of temporal redundancy and the effects of past experience. The learning mechanism, when provided, would appear to be necessarily of the specificity type. At the moment there appears to be no room for new perceptions of the sort considered by the enrichment theorists—notably those of the "new look" variety. This, then, is a theory of literal perception. It is not clear how the scanning mechanism will be instructed to respond when the properties of perception are not directly inherent in the physical object perceived.

III. Recent Physiological Models of Perception

Hebb's cell-assembly theory of perceptual development. Hebb has presented a neurological theory of perception in his *Organization of behavior* (32). While basically of the telephone switchboard type, its design provides the complexity and flexibility necessary to account for the field properties of perception. Its major purpose, however, appears to be to provide a vehicle for explaining what Hebb regards to be the basic role of experience—particularly early experience—in structuring perceptions. Like Berlyne (8), he deals with the phenomena of attention in the same model. Like James (45), he sees even the simpler perceptions in structure going through a long course of development. Like Werner (68), he recognizes the role of motor processes in perception. His model is an elaboration of Lorente de Nó's concept of reciprocal circuits, to which has been added an assumption of axon growth with exercise. (The status of this last assumption has been questioned. Sholl [59, p. 91] cites research by Haddara which reports that while large measurable changes take place in the dendrites of cortical neurons of mice after birth, no difference could be detected between the neuronal development of normal animals and those in which the eyes had been enucleated at birth.) Spread of excitation in the developing nervous system is at first essentially random. With repeated receptor stimulation, however, neurons become organized into functional units called cell assemblies. With continual exercise the stability of these units is enhanced. As experience increases and is enriched, cell assemblies combine into more elaborate temporal organizations called phase-sequences.

With these notions, Hebb explains the configurational properties of perception, expectancy, perceptual learning, perceptual selectivity and perceptual generalization. In short, he accounts for both enrichment and specificity, as well as for the relationship between perceptual and other psychological functions. Hebb's theory is a brilliant intellectual exercise, imaginative and insightful. At present, however, it is difficult to feel that it constitutes any more satisfactory explanatory device than some of the models we have already described. Defining schemata in neurophysiological terms does not make them any more accessible or, for that matter, real. Granted that Hebb has elaborated the internal structure of his model in more specific detail than other theorists, and that he has described in terms of this model the nature of the several psychological functions, he still has to provide a key for transliteration from the terms of his model to the terms of empirical observation. The chemist has operations that enable him both to produce and to identify precise changes in molecular structure. He can talk, for example, of adding a double bond at a certain position in a particular molecule. Can the psychologist talk in the same way about adding a closed-chain circuit to a cell assembly? While it would be manifestly unfair to expect this sort of specificity so soon from Hebb, his theory implies it; and unless a means is clear by which it ultimately may be attained, his picture of central mechanisms, no matter how detailed, is no more than metaphorical.

Werner and Wapner's sensory-tonic field theory of perception. This theory is a recent variant of Gestalt psychology. It begins with a need to account for what its authors refer to as the projective nature of perception as well as its active, interactive characteristic. At the same time Werner and Wapner reject the possibility of a simple motor theory because they feel that such a view could not describe the mechanism that accomplishes sensory-motor interaction (67, 68). Instead, they propose a sensory-tonic theory. Tonus, as they use the term, refers not only to the distribution of muscular tension, but also to the level and pattern of visceral reactivity. Perception is a "total dynamic process" patterned not only by the nature of sensory input but also by the tonic status of the perceiving organism. Sensory and tonic factors not only interact but are also dynamically equivalent.

Werner and Wapner support this last contention by citing evidence from the recent neurophysiological literature that a simple dichotomizing of CNS functions into sensory and motor processes is difficult if not impossible. From the psychological literature they review data that have been traditionally catalogued under the label, "Unity of the senses," and in this connection it should be noted that the tension effects they describe must be mediated through sensory feedback, either visceral or proprioceptive. Their own experimental work demonstrates that the perceived spatial position of simple figures, the autokinetic effect, and the perception of movement in static drawings can be influenced by the asymmetrical arousal of muscular tension (30, 51, 66). For example, the position of the vertical can be shifted left or right by electrically stimulating the neck muscles on the left or right side, respectively. Indeed, though they discuss visceral tensions, and suggest that their theory provides a means of accommodating the problem of motivation and emotion in perception (i.e., the projective nature of perception), their observations have, for the most part, been confined to the relation between tonus in the voluntary musculature and visual orientation. The notion of pooling of effects has already been seen in Helson's adaptation-level theory, and Helson has suggested (34) that his quantitative approach will facilitate an analysis of sensory interaction as well as an understanding of its central mechanism.

Perception in the ethological theory of instinctive behavior. The ethologists, a contemporary group of European zoologists, have elaborated a theory of instinctive behavior that includes some Gestaltist-influenced notions about the nature of perception (61).

The term "instinctive response" covers a varying complexity of animal activities. For example, the simple swimming movements of the eel and the elaborate mating ritual of the stickleback are both typed as instinctive behavior. The appearance of an instinctive response depends upon the activation of an "Innate Releasing Mechanism." The ease with which this mechanism is activated depends in turn upon motivation (physiologically, a concatenation of hormone actions), internal stimulation, and intrinsic nervous activity. Motivation, furthermore, is also susceptible to alteration by certain external conditions called *orienting stimuli.*

Actual arousal of the I.R.M. depends —except in conditions of intense motivation when vacuum activities may occur—upon the animal's being confronted by certain *sign stimuli.* The male stickleback in mating condition must perceive the swollen belly of the female before he executes the courting dance; the young herring gull must perceive the red patch on the mother's bill before giving the begging response; the young gallinaceous bird must perceive the long-necked silhouette of the mother before giving the gaping response. These stimuli vary in complexity. They also have Gestalt character, and cooperate according to what Tinbergen calls the "rule of heterogeneous summation" (61). Thus perceptions, the theory implies, are immediately given, and depend upon species, age, sex, motivational state, etc. Even in the instances where the ethologists recognize learning (they give greatest attention to Prägung, one-trial permanent learning that occurs during some critical period, usually early in the animal's life), this does not include perceptual learning, but rather presupposes, particularly in the case of Prägung, that perceptual organization is immediately achieved. This instinct theory, like all instinct theory, requires the assumption that the effects of learning are confined to the specific area of experience in which learning occurs. Riess' (58) study of "instinctive" nest building, as well as the extensive literature on learning set, raises doubt concerning the validity of such an assumption.

IV. Modern Functionalist Views of Perception

McDougall's conception of perception. The reader may question the appropriateness of including a brief discussion of McDougall's view in this section. But McDougall was among the early protestants against a preoccupation with conscious content. To talk about contents, he insisted, is to lead to continual confusion. Perception, rather, is an active interpretative process of the mind (53). All human beings begin with a capacity for experiencing. After repeated practice in perceptual thinking, all become capable of increasingly refined and complex discrimination and, at the same time, of rich elaboration of the colligation of sensory signs confronting them. Perception is thus a form of conception and it is impossible, as Adams reiterates later (2), to make more than a verbal distinction between them. McDougall's discussion of cognitive dispositions implies a basic unity in the organism's behavior. But there is no suggestion of the dichotomy between specificity and enrichment in perceptual learning that the Gibsons have recently insisted upon (29).

Woodworth's view of perception as motivated response. In the Woodworth of 1921 (72), one notes some clear-cut differences from the conscious-content views described earlier. While some attention is directed toward conscious content, perception is seen primarily as a function performed by certain areas of the brain. It consists of neurological events that follow sensation and precede motor response. This neurologizing in itself did not constitute progress over its contemporary arm-chair analysis clothed in terms of consciousness. But progress did lie in the shift in emphasis from content to function. Perception is a response and not a picture, an active analyzing and interpreting of signals provided in sensory stimulation, a knowing of the facts. As such, it transcends present stimuli and is a necessary antecedent to overt motor response. As an interpretative response it is both combining and isolating in function, the culmination of a process of discovery of information about the world. It is susceptible to the regular laws of learning.

This position is made clearer in Woodworth's 1947 note on conditioning (73). What is reinforced, he contends, is the perceptual response—something on the receptor rather than the effector side. Certain views of the world get fixed as a result of the consequences they have for the organism's adjustment to its environment. Thus the common impression that perception is a passive affair derives from the fact that certain views provoke almost immediate reaction. But when a new perception is in the making, it is seen to be a slow trial-and-check affair motivated by a basic will to perceive.

Here, then, perception is not a picture of the world, but a motivated process, part of the organism's active search for adjustment in the world. Woodworth's view of perceptual learning resembles, in at least one aspect, that of James: the result of continued reinforcement must certainly be "old-fogeyism." The description of perception as an active process identifies Woodworth with the transactionalism of the functionalists of the 1950's. But in Woodworth, as in James, the notion of a reciprocal relationship between perception and action is not explicit. While perception as an antecedent to motor response most certainly influences the latter's character, the alternative possibility is not made clear.

Brunswik's functionalism. Brunswik may be classed among the modern functionalists in perception. He has left us, however, no highly elaborated theoretical structure, for his preoccupation was with methodological issues. It was his conviction that the nature and magnitude of S-R correlations must be established before a theoretical superstructure can be safely put up. His functionalism is indicated at the level of theory construction by his concept of *ecological validity,* and at the level of methodological implementation by his notions of *representative design.* It is also seen in the correlational character of his work and in his interest in probability learning—particularly in his work on perceptual probability learning (18). Here it is clear that he meant to relate the very structure of perception to pattern of past experience. His work on distal focusing and the constancies (16) reveals a bias toward veridicality, but this he correctly sees as a matter of degree, not as one side of a theoretical issue. Brunswik's view is relational-organizational in character. His views on probability learning and his concern with the relation of perception to goals and aims resemble aspects of James and of Helmholtz. There is an obvious similarity between what takes place in probability learning and in the development of Bruner and Postman's families of hypotheses (13) and of Cantril's assumptive world (21), though Brunswik's notions are probably more precisely formulated. Finally, we may suggest that probability learning is a quantitative complement of adaptation-level.

The transactional approach of Ames and Cantril. The view evolved by Ames, Cantril, and their students (21, 42) contains elements similar in some respects to a number of the views previously described—e.g., James (45), Brunswik (17), Woodworth (73), Hebb (32), and Helson (34). Indebtedness to Whitehead (70) and to Dewey and Bentley (23) is expressly recognized.

Perception, according to the transactionists, is a process by which an individual person, with a particular outlook, attributes significance to events in his behavioral environment. These significances are established through the past attainment of goals. Put more simply, perception is part of the process of interpreting present stimulus cues from the bias of expectation. Expectations, or assumptions, are the consequents of past experience and action. As they are repeatedly confirmed or reinforced (and it is appropriate to speak of reinforcement, since the criteria of evaluation are provided by the purposes of the individual perceiver), corresponding facets of perception become more stable and harder to alter—although changes within specific situations may be dramatically effected by a single failure of confirmation.

While there is no detailed treatment of the generality of an assumption and its strength in a particular situation, one may expect that persistent failure in confirmation in some particular situation may result in weakening of the assumption for other situations. There is, this group points out, almost an infinity of physical arrangements (equivalent configurations) which will produce the same pattern on the retina (43). These are thus all seen alike, and what is seen is what one has been familiar with through the transaction of living. Rooms may be constructed in a variety of physical shapes. All will have the same perceived shape—the cubicle of everyday experience. This perception can change dramatically, however, when the viewer checks his assumptions about shape by feeling about the room and finding it not to be of normal shape. Presumably, if failure of confirmation of the assumption of cubeness occurred frequently enough, our normal response to the equivalent configurations for rooms would also change. The totality of assumption, strong and weak, that the perceiver holds about his surroundings constitutes his assumptive world. The physical world must thus remain an inference. Perceiving, not a passive imprinting of a world picture, is one aspect of an active transaction between the striving organism and its environment. Perception, therefore, is necessarily the product of learning. It is not immediately given by the properties of the physical world and receptor organ. The demonstrations devised by the transactionists, when considered together, emphasize the lack of correspondence between the physical and phenomenal worlds: perception is nonveridical, but it is also stable. It is, of course, not exactly appropriate to speak of nonveridical perception, since we are not dealing with a noncorrespondence between the real world and one's impression of it, but rather with noncorrespondence between inferences about the world described in phenomenological and in physical terms.

The transaction theory has been criticized for ignoring the possible role of stimulus configuration in the structuring of perception (27). But neither the theorists nor their critics have created a crucial situation for establishing the primacy of either stimulus configuration or learning. It has also been contended that the transactional view fails to provide for perceptual learning of the specificity type. This may be countered by pointing out that strengthening of assumptions is a matter of differential reinforcement, and it is exactly with this technique that specificity is established (e.g., in the conditioned salivary response of the dog to tones). When strength of hypothesis is defined in terms of probability statements, we note the similarity of this view to that of Brunswik. When the probabilities are described as weighted in terms of relevance, the possibility of formulation of hypotheses in terms of norms, in the manner of Helson's adaptation-level, becomes apparent. Specificity is then definable in terms of variable error about this norm.

Bruner and Postman and the "new look" in perception. The late 1940's and early 1950's saw a long and controversial series of experiments directed toward testing the functionalist proposition that perception is an instrumental activity serving the needs and wants of the individual perceiver. Bruner and Postman were leaders in this movement. It was their conviction that we perceive what we need to perceive and what we are used to perceiving (14). Autochthonous factors were accepted as a class of perceptual determinants, and it was recognized that there are situations where the properties of the stimulus configuration leave nothing more to be determined in perception. But it was maintained that autism in perception typically is strong. For example, when it was argued that the "new look" effects are produced in the laboratory under conditions of extreme stimulus impoverishment, the reply was that the conditions of viewing in everyday life (e.g., lack of specific expectation, the momentary glance, etc.) are conditions of stimulus impoverishment. The motivated perceiver is alert to the presence of personally relevant objects and, although on certain occasions recognition of their presence may be withheld from consciousness, he recognizes them more readily, tends to overestimate their size, etc. Thus accentuation, fixation, vigilance, defense, resonance, and selectivity were presented as explanatory principles to take their place along with closure, symmetry, proximity, similarity, and good continuation.

The earliest attacks on the "new look" approach were criticisms of technique. Failures at replication were described, and flaws in experimental design pointed out (54). But the theoretical formulation has also been attacked as vague and equivocal. Luchins (52) rejects the distinction between formalism and functionalism, marshaling a great mass of data that demonstrate perception to be stable and veridical, and insists that the importance of rational processes in perception has been neglected. Hochberg and Gleitman (40) object that the Bruner-Postman type of theorizing makes explanations possible simply by elaborating, *post hoc,* the list of motives presumed to underlie perception. Howie (41) claims that the theory does not enable one to determine when one functional principle gives way to another, and Prentice (57) suggests that to define perception in the way that Bruner and Postman do is not only to inflate it beyond its traditional meaning but to reduce it to meaninglessness.

Bruner and Postman have come to recognize their principles (perceptual defense, accentuation, etc.) as descriptive rather than explanatory (13, 56), and have proposed a more general expectancy-type theory. According to this later formulation, perception begins with expectancies or *hypotheses.* Each hypothesis represents the arousal of a complex of cognitive and motivational processes by an environmental clue. The present stimulus situation provides "information" to the expectant organism. This is evaluated with expectation as the criterion. If the initial hypothesis is confirmed in some critical degree, it is strengthened; if not, it is set aside for a new hypothesis. Hypothesis strength depends on frequency of past confirmation, the number of alternative hypotheses available, and certain cognitive, motivational and social consequences. Like Berlyne, Bruner and Postman attempt to account for the active, directive characteristic of perception by proposing a mechanism that unites its "pictorial" aspects with set. If one thinks of hypothesis strength in terms of relative frequency of reinforcement, one recognizes the striking similarity between this view and that of the transactionist. Meanwhile, while expectancy is the basic functional principle, it is not clear how it relates, within the framework of the theory, to other psychological processes like the mechanisms of specific motives and of the several cognitive processes, e.g., memory. Distinction between hypotheses and other psychological processes at the operational level in specific instances of observation appears to be momentarily extremely difficult. Brunswik (17), meanwhile, has made a definitional distinction between perception and thought in terms of uncertainty-geared and certainty-geared functions. Operational distinction was demonstrated by presenting a size-constancy task like, on the one hand, a problem in direct perceptional observation, and, on the other, an arithmetic-reasoning problem. In the first case, responses were compactly and fairly normally distributed, and indicated approximate size constancy. In the latter case, there were many on-the-dot correct responses, but also many extreme and bizarre errors indicating confusion and the shortcomings of logical extension from misinterpreted cues.

Vernon's application of the concept of schema. Vernon's theory of perception is a functional theory that stresses the stability and orderliness of the perceiver's conception of the world (65). She is critical of the recent trend toward studies of motivation and perception for ignoring the intelligent operation of the perceiver's background of knowledge in structuring perceptions, and for overestimating the transient effects of needs. She begins by recognizing that the external world consists of orderly arrangements of objects, and that these give rise to a stimulation pattern that is ordered. Meanwhile, she rejects Gibson's asumption (26) that ordinal stimulation gives immediate rise to the finished perceptual product. It is here believed that perceptions are the end products of a slow process of construction through manipulation of and experimentation with the external world. While admitting that the slow perceptual development suggested by Piaget (e.g., 55) may, to some extent, be attributable to maturation, she insists that specific experience is essential to the perceiver's conception of himself and the world. Through experience is evolved a system for integrating and interpreting the information provided in stimulus input. Motivation plays a major role in both the differentiation and elaboration of these classificatory criteria.

This theory is, of course, an application of Bartlett's concept of *schema* to the process of perceiving (7). Its resemblance to Bruner and Postman's cycle of hypothesis formation, hypothesis evaluation, and hypothesis reformulation is clear. Unlike Bruner and Postman's theory, however, it stresses the long-term orderliness and veridicality of perception. Long-term orderliness is also seen in the transactionists' notion of the assumptive world, but the insistence upon veridicality is missing. Furthermore, to make the notion of *schema* comparable to the concept of *assumptive world*, it would be necessary to postulate explicitly some superordinate schema that would integrate the various schematic structures (e.g., those having to do with temporal and spatial properties, those dealing with their significances, etc.) into a single functional system. The long time necessary for the elaboration of a schema suggests that its physiological counterpart might be Hebb's cell assembly.

While Vernon has taken pains to discuss the relation of schemata to both perception and motivation, she has not made quite clear the properties of schemata, defined either in terms of details of theoretical structure or operations which differentiate these details at the empirical level. How are schemata differentiated from perceptions, memories, motivations? How are some schemata distinguished from others? How are they functionally related? What are the functional properties of schemata—their rigidity, scope, complexity, etc.? Until questions like these can be answered, we have achieved only a way of looking at perceptual phenomena, not a theory of perception.

Allport's concept of structure. In his recent book (4), Floyd Allport evaluates some 13 perceptual theories and formalizes his own views. In all theories reviewed he finds two common limitations: ineffective handling of the problem of meaning, and unwarranted preoccupation with quantitative law. More basic than this latter, he contends, is another type of law, that of event-structure.

Structures are defined through a geometry of on-goings and their interrelations at event-points. Time is defined, not as a linear function, but as the duration of events that close themselves through a cycle of operations. On-goings are activities distributed both spatially and temporally; event-points are points of "contact" among "on-goings." Structures are not randomly organized patterns, but may be predicted in probability terms. Events, since they are points of contact, have no dimensions; they simply occur. Hence loci within a structure are event-regions of space through time in which events may or may not occur.

Perceptions are dynamically operating structures. Perceptual aggregates resemble Hebb's cell-assembly-phase-sequence scheme (32), save that Allport appears to push the notion beyond the microscopic level of the material hypothetical construct to the submicroscopic level of the nonmaterial, purely intervening variable. Should the theorist be able to formalize structures in unambiguous denotable terms, he would be able to achieve a fruitful, integrated theory. This is an aim that the present writer does not see that Allport—despite compulsive elaborateness and sometimes excruciating logic—has attained. One comes away from reading Allport with the feeling that he has been concerned more with translating the concepts of competing perceptual theories into the idiom of his own viewpoint and less with the construction of a set of sentences that will provide for the explanation and prediction of a repertoire of specific perceptual phenomena. When, for example, the constancies are discussed, it is in terms of finding a basis for Brunswik's probabilistic weighting of cues in the density of meaning-cycles and in interstructurance. When phenomena are analyzed (e.g., thirst-quenching behavior), one is likely to feel that Allport has done little more again than translate the elements of the observational sequence into the nomenclature of his hoop (cycle) analogy. In short, his perceptual theory (and there is none distinct from general structure theory which is applicable to all natural events) is so abstract and is stated in such broad terms that it contributes no more than a general bias—granted that this may eventually prove quite fruitful. Certainly it appears to lack the articulation needed in a theory that is going to be intimately tied to the work-a-day activities of the laboratory.

THE LOGICAL STATUS OF CURRENT PERCEPTUAL THEORY

Early conceptualizations of perception are marked by a concern with conscious content; later views regard perception as a class of response, different from other classes in accessibility (it is inferred rather than observed) and in function (it is an intermediary rather than a consummatory response). Perception as intermediary response was first assumed to stand in a unilateral relation to consummatory acts; more recently the relationship has been described as interactive. Definitions have been framed in phenomenological, physiological, and psychophysiologically neutral terms. Perceptions have, on the one hand, been equated with simple discriminations, and, on the other, identified as events so complex as to defy ready replication. Early use of the term included reference to both the physical properties of an object and its significance; later, primary concern was with object qua object; currently interest has again turned to the meaningful object. Earlier studies dealt primarily with the encyclopedic task of analyzing and cataloging the properties of perception, generically defined; modern studies are most frequently directed toward schematizing the interplay of factors that give rise to some particular set of perceptual responses. While some current work stresses the autistic properties of perception, most of it emphasizes veridicality. Early theorists regarded learning to be of paramount importance in determining perceptual structure; the Gestalt theorists of the middle period de-emphasized the importance of learning, with their powerful demonstrations of the organizing factors; modern theorists have restored learning, particularly early learning, to a position of prominence. Except in modern psychophysics, attention has been currently directed toward independent variables of a central rather than peripheral sort.

The adequacy of current data for theory construction. The adequacy of available perceptual data for the construction of a general theory of perception has been questioned, for one reason or another. A common criticism leveled against early work was that the criteria for identifying the perceptual response were basically individual and inside the observer. This actually meant that the criteria for identifying the data lacked precision, and that the status of the data with regard to the requirement of intersubjectivity was undetermined. Of course, all perceptions have in common the fact that they can be made public only as discriminations, and, unfortunately, the methodological reaction against introspection has resulted, in most cases, in a highly restricted definition of discrimination. Until recently, experimental designs have usually allowed for only a two-category response with all but one of the dimensions that might affect the response held constant. While the objectivity of the resulting data is beyond question, the value of such designs for the creation of the multidimensional interactive theories currently favored is unclear—such designs frequently lack representativeness (17). But before we can achieve data commensurate with such a view, perception must be dimensionalized. The structure of the underlying mechanism cannot be defined fully by studies that provide for only the identification of relevant variables, although this is a necessary first step. This is tantamount to saying that the psychophysical approach must be exploited, regardless of one's theoretical bias, before higher-order theories can be competently achieved.

Perception—a theoretical concept. It is not improbable that the long-time preoccupation with conscious content resulted from a failure to recognize the distinction between perceptual phenomena and perception as a theoretical concept. But as Tolman says (63, p. 111), "The fundamental essence of any hypothesis is not, it would seem, whether it be conscious or not but the fact that it is a condition in the organism which sets the latter, whether human or subhuman, for a certain systematic selectivity of behavior."

Logically, perception is a dispositional concept, and, as indicated by the occurrence of controversies like nativism versus empiricism, organization versus stimulus-compounding, and autism versus veridicality, a dispositional concept cannot be defined by *any* fixed totality of reduction sentences. It is always possible that more will be added as observational situations change. In the language of the experimenter, these reduction sentences constitute a class of operational definitions. Hence it further follows, as Garner, Hake and Eriksen (25), and Bridgman (12) before them, have pointed out, no concept can be given definitive definition by a single operation. Overlooking this fundamental requirement has led many investigators to the misimpression that operational validity is, in the final sense, simply a matter of arbitrary definition. Thus perception has been equated with simple discrimination. This view has yielded clear-cut experimental results, but under such restricted conditions that their generality is unclear. Further, it has meant an inability to differentiate between the properties of a perceptual mechanism which we infer to exist and those of receptor and effector systems. Thus distinction between sensation and perception is a priori, and made purely at the semantic level. (It is recognized that there are levels of communication—i.e., in an achievement-oriented psychology like Skinner's with the white rat—where the distinction between central and peripheral mechanism is not necessary.) Garner *et al.* (25) cite perceptual defense as a recently promulgated concept that is ambiguous because the experiments from which it was derived failed to achieve operational clarity. But lack of clarity probably results also from a lack of appreciation of the function that the concept is to perform. (See, for example, Gibson's paper, "What is Form?" [28], in which he describes 10 different concepts of form encompassed in three general categories of definition.) The remarks of the present paragraph are not intended to imply that perceptions should not be identified with discrimination for, as was pointed out earlier, discriminatory responses are the only means by which perceptions are made accessible as objective data. But the concept, as it has evolved from human studies, implies multidimensionality. Thus, it requires objective definition by a system of discriminations, as Boring suggests (11), made explicit through experimentation, if only such experiments are performed.

A general weakness of perceptual theories. Perception as a dispositional concept depends as much upon assumptions associated with the invented mechanism as upon the empirical data to which it is operationally linked.

Perhaps the most widespread methodological defect of perceptual theories (and this holds for psychological theories in general) is the absence of a set of statements that relate propositions about the model to reduction sentences (what Campbell picturesquely called the "dictionary" [20]). Even where the model (e.g., Hebb's cell-assembly model) is delineated in detail, these are more or less lacking. Perhaps a prominent exception is Köhler and Wallach's formulation of the satiation theory of figural aftereffects. Failure to relate data systematically to characteristics of the mechanism has meant an inability to relate various properties of the mechanism to each other. Thus, it is difficult to make a scientifically meaningful distinction between sensation and perception, and perception and conception, or to differentiate operationally between the processes of attention and those of perception. Unless such theoretically grounded, operationally specified distinctions can be made there is no justification for these terms in the formal language of psychology.

Prospectus. Now that charges have been drawn, what must be done to improve the status of perceptual theory? Armed with the equipment of modern logical analysis, one might attempt a tour de force, creating a highly articulated theoretical structure, cataloging and correlating reduction sentences, placing assumptions about the model in propositional form, and identifying the conditional reduction sentences necessary to convert propositions into statements about data. Such an exercise would probably not come off. Science, as Conant points out (22), is not made that way. Unlike a system of formal logic, it is not closed. Rather it is always in the status of becoming—of uncovering new data and incorporating new concepts.

It may be advantageous, however, to explore our dependent variables, both intervening and otherwise, as dimensions to be scaled. Only after this is achieved can the characteristics of our data be tied to the structure of our model thoroughly enough to provide high-level explanations. My personal preference at this stage of psychology's development is for formal constructs. These do not commit us to a physiology that may some day trap us. (As Adams [3] insists, a psychical system may be just as objective as a reflex arc or a cell assembly.) They can be conceived of, first, in simple terms with few dimensions, and expanded and made more complex as data are acquired, without risking the necessity of having some day to be totally abandoned in the face of some impressive piece of incompatible evidence. And finally, their properties are easily conceived of as dimensions to be scaled. The possible fruitfulness of such an approach is indicated by Helson's work on adaptation level.

REFERENCES

1. ADAMS, D. K. The inference of mind. *Psychol. Rev.*, 1928, **35**, 235–252.
2. ADAMS, D. K. *The anatomy of personality.* New York: Doubleday, 1954.
3. ADAMS, D. K. Learning and explanation. In *Learning, personality theory and clinical research, the Kentucky Symposium.* New York: Wiley, 1954. Pp. 66–80.
4. ALLPORT, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
5. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.
6. ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.
7. BARTLETT, F. C. *Remembering.* Cambridge: Cambridge Univer. Press, 1932.
8. BERLYNE, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
9. BEVAN, W., JR. Modern psychologists: scientific woozle hunters? an opinion in outline. *Nordisk Psychologi's Monografiserie.* NR. 4, 1953, 1–28.
10. BEVAN, W., & DARBY, C. L. Patterns of experience and the constancy of an indifference point for perceived weight. *Amer. J. Psychol.*, 1955, **68**, 575–584.
11. BORING, E. G. The use of operational definitions in science. *Psychol. Rev.*, 1945, **52**, 243–245.
12. BRIDGMAN, P. W. *The logic of modern physics.* New York: Macmillan, 1927.
13. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey (Eds.), *Perception, an approach to personality.* New York: Ronald, 1951. Pp. 121–147.
14. BRUNER, J., & POSTMAN, L. Perception, cognition and behavior. *J. Pers.*, 1949, **18**, 14–31.
15. BRUNSWIK, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
16. BRUNSWIK, E. Distal focussing of perception: size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56**, No. 1 (Whole No. 254).
17. BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley, Calif.: Univer. Calif. Press, 1956.
18. BRUNSWIK, E., & HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, **41**, 281–290.
19. BRUNSWIK, E., & KAMIYA, J. Ecological cue-validity of "proximity" and of other Gestalt factors. *Amer. J. Psychol.*, 1953, **66**, 20–32.
20. CAMPBELL, N. R. *Physics: the elements.* Cambridge: Cambridge Univer. Press, 1920. Pp. 120–140.
21. CANTRIL, H. *The why of man's experience.* New York: Macmillan, 1950.
22. CONANT, J. B. *On understanding science.* New Haven: Yale Univer. Press, 1947.
23. DEWEY, J., & BENTLEY, A. F. *Knowing and the known.* Boston: Beacon Press, 1949.
24. DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).
25. GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.
26. GIBSON, J. J. *The perception of the visual world.* New York: Houghton-Mifflin, 1950.
27. GIBSON, J. J. Theories of perception. In *Current trends in psychological theory.* Pittsburgh: Univer. Pittsburgh Press, 1951. Pp. 85–110.
28. GIBSON, J. J. What is form? *Psychol. Rev.*, 1951, **58**, 403–412.
29. GIBSON, J. J., & GIBSON, E. J. Perceptual learning: differentiation or enrichment. *Psychol. Rev.*, 1955, **62**, 32–41.
30. GOLDMAN, A. E. Studies in vicariousness: degree of motor activity and the autokinetic phenomenon. *Amer. J. Psychol.*, 1953, **66**, 613–617.
31. GRAHAM, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.
32. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
33. HELSON, H. Adaptation-level as frame of reference for prediction of psychophysi-

cal data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

34. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **58**, 297–313.

35. HELSON, H. Psychiatric screening of flying personnel. Perception and personality—a critique of recent experimental literature. U.S.A.F. Sch. Aviat. Med. Prog. Rep. No. 21-0202-0007, Rep. 1, July 1953.

36. HELSON, H., BLAKE, R. R., MOUTON, J. S., & OLMSTEAD, J. A. Attitudes as adjustments to stimulus, background, and residual factors. *J. abnorm. soc. Psychol.*, 1956, **52**, 314–322.

37. HELSON, H., DWORKIN, R. S., & MICHELS, W. C. Quantitative denotations of common terms as a function of background. *Amer. J. Psychol.*, 1956, **69**, 194–208.

38. HELSON, H., JUDD, D. B., & WILSON, M. Color rendition with fluorescent sources of illumination. *Illum. Eng.*, 1956, **51**, 329–346.

39. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.

40. HOCHBERG, J. E., & GLEITMAN, H. Toward a reformulation of the perception-motivation dichotomy. *J. Pers.*, 1949, **18**, 180–191.

41. HOWIE, D. Perceptual defense. *Psychol. Rev.*, 1952, **59**, 308–315.

42. ITTELSON, W. H., & CANTRIL, H. *Perception, a transactional approach.* New York: Doubleday, 1954.

43. ITTELSON, W. H., & KILPATRICK, F. P. Equivalent configurations and the monocular and binocular distorted rooms. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Inst. for Associated Res., 1953. Ch. 3.

44. JAFFE, R. The influence of visual stimulation on kinesthetic figural after-effects. *Amer. J. Psychol.*, 1956, **69**, 70–75.

45. JAMES, W. *Psychology, briefer course.* New York: Holt, 1904.

46. KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt Brace, 1935.

47. KÖHLER, W. Bemerkungen zur Gestalttheorie (in Anschluss an Rignanos Kritik). *Psychol. Forsch.*, 1928, **11**, 188–234.

48. KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1925.

49. KÖHLER, W., & WALLACH, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

50. KÖHLER, W., HELD, R., & O'CONNELL, D. N. An investigation of cortical currents. *Proc. Amer. Phil. Soc.*, 1952, **96**, 290–330.

51. KRUS, D. M., WERNER, H., & WAPNER, S. Studies in vicariousness: motor activity and perceived movement. *Amer. J. Psychol.*, 1953, **66**, 603–608.

52. LUCHINS, A. An evaluation of some current criticisms of Gestalt psychological work on perception. *Psychol. Rev.*, 1951, **58**, 69–95.

53. MCDOUGALL, W. *Outline of psychology.* New York: Scribners, 1923.

54. PASTORE, N. Need as a determinant of perception. *J. Psychol.*, 1949, **28**, 457–476.

55. PIAGET, J. La construction du réel chez l'enfant. Neuchatel: Delachaux et Niestlé, 1937.

56. POSTMAN, L. Toward a general theory of cognition. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harpers, 1951. Pp. 242–272.

57. PRENTICE, W. C. H. Functionalism in perception. *Psychol. Rev.*, 1956, **63**, 29–38.

58. RIESS, B. F. The isolation of factors of learning and native behavior in field and laboratory studies. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1093–1102.

59. SHOLL, D. A. *The organization of the cerebral cortex.* New York: Wiley, 1956.

60. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953. Pp. 257–282.

61. TINBERGEN, N. *Instinct.* Oxford: Clarendon Press, 1951. Ch. 2–5.

62. TITCHENER, E. B. *A textbook of psychology.* New York: Macmillan, 1910.

63. TOLMAN, E. C. *Collected papers in psychology.* Berkeley, Calif.: Univer. Calif. Press, 1951.

64. TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–327.

65. VERNON, M. D. The functions of schemata in perceiving. *Psychol. Rev.*, 1955, **62**, 180–192.

66. WAPNER, S., & WERNER, H. Gestalt laws of organization and organismic theory of perception: effect of asymmetry induced by the factor of similarity on the position of the apparent median

plane and apparent horizon. *Amer. J. Psychol.*, 1955, **68**, 258–265.

67. WERNER, H., & WAPNER, S. Sensory-tonic field theory of perception. In J. S. Bruner and D. Krech (Eds.), *Perception and personality, a symposium.* Durham, N. C.: Duke Univer. Press, 1950. Pp. 88–107.
68. WERNER, H., & WAPNER, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–338.
69. WERTHEIMER, M. Untersuchungen zur Lehr von der Gestalt, II. *Psychol. Forsch.*, 1923, **4**, 301–350.
70. WHITEHEAD, A. N. *The aims of education.* New York: Mentor Books, 1949.
71. WICKENS, D. D. Stimulus-response theory as applied to perception. In *Learning, personality theory and clinical research, the Kentucky Symposium.* New York: Wiley, 1954. Pp. 22–35.
72. WOODWORTH, R. S. *Psychology, a study of mental life.* New York: Holt, 1921.
73. WOODWORTH, R. S. Reënforcement of perception. *Amer. J. Psychol.*, 1947, **60**, 119–124.

(Received April 22, 1957)

Psychological Review
Vol. 65, No. 1, 1958

A SIMPLIFIED MODEL FOR STIMULUS DISCRIMINATION [1]

EDWARD J. GREEN

Dartmouth College

It has frequently been noted during the early phases of discrimination training that response probability increases in the presence of negative instances, or S^{Δ}'s. That is, responding in the presence of such stimuli does not decrease monotonically in strength as is the case in simple extinction. Skinner has attributed the temporary increase in S^{Δ} response strength to induction (**14**). Induction is defined as a consequence of there being common stimulus properties between S^{D}'s and S^{Δ}'s.[2] The greater the similarity between the stimulus sets, the more induction takes place. The more induction present during discrimination training, the more difficult is the establishment of a sharp discrimination.

One of the major contributions of the Bush and Mosteller paper on stimulus generalization and discrimination (**2**) is the formalization of the properties of the discrimination situation. To recapitulate somewhat, they categorize subsets of stimulus elements as shown in Fig. 1, where S is the set of stimulus elements which constitute the occasion for reinforcement if a response occurs. S' is the set which is the occasion for nonreinforcement if the organism responds. I is the intercept of S and S'. It should be noted that I exists *always* in discrimination, whether or not the experimenter so contrives it. For example, the intercept includes all uncontrolled stimulation common to both S and S' stimulus sets, in addition to whatever the experimenter may or may not have contrived experimentally to be common to both sets.

Certain objections can be raised to the Bush-Mosteller model in regard to the number of constants to be estimated, and in particular to the statement that the measure of the intercept decreases as discrimination learning increases. Clearly the physical nature of the stimulus set is unchanged by the fact that an organism has come to make a discrimination. The present development owes much to the Bush-Mosteller formalization, but takes its roots in the independent though par-

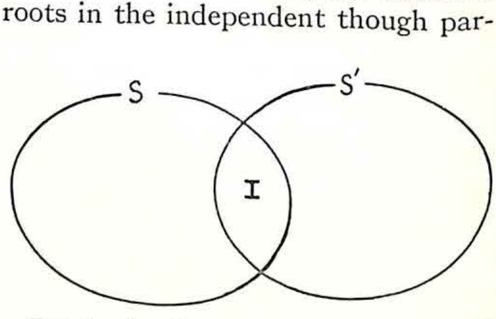


FIG. 1. Set diagram indicating overlap (I) between stimulus elements present on occasions for reinforcement (S) and elements present on occasions for nonreinforcement (S').

---

[1] The writer is greatly indebted to J. L. Snell and J. G. Kemeny of the Department of Mathematics at Dartmouth College for assistance and encouragement in the development of the model.

[2] Two conventions will be used throughout the remainder of this paper. In place of S^{D} and S^{Δ}, S and S' will be employed respectively to facilitate identification of parameters in this and the Bush-Mosteller paper. In addition, in the treatment of discrimination learning, the term *trial*, as employed by Estes *et al.*, is used in a broad sense to include the presentation of discriminative stimuli in a series. Rate of response has been shown to be directly related to probability of response as treated herein, and this usage makes evident certain possible extensions of the model to the free operant situation.

allel learning model of Estes and Burke (5). As in both the Bush-Mosteller and the Estes-Burke models, conditioning and extinction are conceived of simply as probability transfers between response classes. In this treatment the additional simplifying assumption is also made that behavior other than that under immediate consideration is mutually exclusive of that behavior. In simple conditioning the mean probability of response on the nth trial has been defined to be:

$$\bar{p}_n = 1 - (1 - \bar{p}_0)(1 - \theta)^n. \quad (1)$$

Symmetrically, in extinction, probability of response is:

$$\bar{p}_n = \bar{p}_0(1 - \theta)^n, \qquad (2)$$

where $\bar{p}_0$ is the initial probability of response, or operant level; θ is the sampling ratio, or the probability of sampling a particular element; and n denotes the number of trials.

THE MODEL

Basically, one may conceive of discrimination as being the simultaneous or successive (depending upon whether stimuli are presented in pairs or serially) conditionings and extinctions of behavior appropriate to the sets of stimuli employed. The present treatment deals primarily with serial presentation, but can be adapted to the simultaneous procedure. It is appropriate to observe that Equation 1 describes the expected change in response probability associated with stimulus subset $S-I$, and that Equation 2 describes theoretical changes in behavior associated with $S'-I$. Were it not for the intercept elements, the present model would simply predict symmetrical changes in S and S', but the behavioral dynamics associated with I complicate the picture.

Estes and Straughan (6) developed a more generalized function to describe the results of sampling from the same stimulus set in both conditioning and extinction trials. The function is:

$$\bar{p}_n = \pi - (\pi - \bar{p}_0)(1 - \theta)^n, \quad (3)$$

where π represents the relative frequency of reinforced trials in the stimulus series. One may regard this function as representing a special case of discrimination, where probability of sampling a stimulus element is the same whether or not it is in S or in S', and where discrimination between S and S' is impossible. Within this framework, π is the asymptote which response probability approaches.

θ is defined as the probability of sampling a particular stimulus element. In the original paper by Estes (3), the sampling ratio was also defined as the ratio s/S, where s represented the number of elements sampled on a given occasion, and S represented the size of the population to be sampled. It is assumed that the *number* of elements sampled on a given occasion, s, remains constant. With θ so defined, it is seen to be determined both by organismic and by environmental factors. For example, should the organism sample exactly one element on any given trial, and should the population to be sampled consist of exactly one element, then $\theta = 1.00$. If, on the other hand, the population should consist of twenty elements, the value of θ would be 0.05. Thus θ is seen to depend upon environmental factors. Should the organism suffer impairment of its discriminatory apparatus, one might predict that the number of elements which it is capable of sampling on a given occasion is reduced. Thus the effect upon the value of the sampling ratio of impairing the discriminatory apparatus would be comparable to increasing the size of the stimulus population. Burke, Estes, and Hellyer (1) have shown that rate of conditioning is inversely proportional to signal variability. Increasing signal variability, in other words, is one way to decrease the sampling ratio. Similar observations have been made in a more conventional discrimination experiment (8).

Once it is assumed that θ is a parameter determined both by organismic and by environmental factors, one may examine the consequences of assigning different values to θ under conditions of reinforcement and nonreinforcement. Let $\bar{\pi} = 1 - \pi$, θ_1 being the sampling ratio for S elements, and θ_2 the sampling ratio for S' elements.

Consider a single element in the intercept. One can form a Markov chain with States 0 and 1. State 0 means that the element is not conditioned, and 1 that it is (11, Chap. 7, Sec. 5). The transition matrix for this chain is:

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} \quad 0 \quad & \quad 1 \quad \end{array} \\ \begin{bmatrix} 1 - \pi\theta_1 & \pi\theta_1 \\ \bar{\pi}\theta_2 & 1 - \bar{\pi}\theta_2 \end{bmatrix} \end{array}$$

The entry $p_{ij}^{(n)}$ of P^n gives the probability that the system is in State j after n steps if it starts in State i. But

$$P^n = \frac{1}{\pi\theta_1 + \bar{\pi}\theta_2} \times \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} \qquad 0 \qquad & \qquad 1 \qquad \end{array} \\ \begin{bmatrix} \bar{\pi}\theta_2 - \pi\theta_1\lambda^n, & \pi\theta_1 - \pi\theta_1\lambda^n \\ \bar{\pi}\theta_2 - \bar{\pi}\theta_2\lambda^n, & \pi\theta_1 + \bar{\pi}\theta_2\lambda^n \end{bmatrix} \end{array}$$

where $\lambda = (1 - \pi\theta_1 - \bar{\pi}\theta_2)$ is the principal eigenvector for P.

The mean response probability with a single stimulus element is $p_{01}^{(n)}$ if the element was initially in State 0, and $p_{11}^{(n)}$ if it was in State 1. Assume now that there are N stimulus elements in the intercept, a fraction, $\bar{p}_0$, of which are conditioned; i.e., in State 1. Then the mean probability of a response becomes:

$$\bar{p}_n = \bar{p}_0 \cdot p_{11}^{(n)} + (1 - \bar{p}_0) p_{01}^{(n)} \qquad (4)$$

which is

$$\bar{p}_n = \frac{\pi\theta_1}{\pi\theta_1 + \bar{\pi}\theta_2} - \left[\frac{\pi\theta_1}{\pi\theta_1 + \bar{\pi}\theta_2} - \bar{p}_0 \right] \times (1 - \pi\theta_1 - \bar{\pi}\theta_2)^n. \qquad (5)$$

It can be seen that this last expression reduces to Equation 3 when $\theta_1 = \theta_2$. It follows that probability of response in S is a weighted mean of Equations 1 and 5 if we somehow take into consideration the fact that the intercept elements are presented within every stimulus presentation in the series, and $S-I$ and $S'-I$ are presented with frequency π and $\bar{\pi}$, respectively. By altering Equation 1 to

$$\bar{p}_n = 1 - (1 - \bar{p}_0)(1 - \theta_1)^{\pi n}, \qquad (6)$$

and Equation 2 to

$$\bar{p}_n = \bar{p}_0 (1 - \theta_2)^{\bar{\pi} n}, \qquad (7)$$

one accomplishes the necessary correction. Now $\bar{p}_n$ refers to a response probability in the presence of the nth stimulus, regardless of whether or not it is in S or S'. Where π equals $\bar{\pi}$, I is operated upon twice as often as either $S-I$ or $S'-I$. If we define $\bar{p}_n(S-I)$ to be $\bar{p}_n$ of Equation 6, and let $\bar{p}_n(I)$ be $\bar{p}_n$ as in Equation 5, then

$$\bar{p}_n(S) = k\bar{p}_n(S-I) + (1-k)\bar{p}_n(I) \qquad (8)$$

and, if we let $\bar{p}_n(S'-I)$ be $\bar{p}_n$ as in Equation 7, then

$$\bar{p}_n(S') = k'\bar{p}_n(S'-I) + (1-k')\bar{p}_n(I) \qquad (9)$$

where k denotes the proportion of elements in $S-I$, k' the proportion of elements in $S'-I$, and $(1-k)$ and $(1-k')$ the appropriate proportions of elements in I for S and S', respectively.

PROPERTIES OF THE MODEL

The present model predicts the characteristic increase and decrease in S' response probabilities under certain special conditions. If $\bar{p}_0$ is greater than 0.00 and less than the asymptotic probability for intercept elements, and if the total number of elements within the intercept constitutes an appreciable proportion of the total in S', then the model generates the S' peak. The increase in S' response probability is produced when the conditioning to the intercept elements is more rapid than is the extinction to the $S'-I$ elements. Once the asymptote for intercept elements is approached closely enough and the rate of conditioning decreases, then extinction to $S'-I$ elements becomes evident, and S' response probability appears to reverse and decline to its terminal value. The model predicts that no initial increase would result if $\bar{p}_0$ were equal to or greater than $\dfrac{\pi\theta_1}{\pi\theta_1+\bar{\pi}\theta_2}$, since $\bar{p}_n(S')$ would decrease monotonically. Likewise, if $\bar{p}_0$ were 0.00, $\bar{p}_n(S')$ would rise monotonically to some terminal value.

Since

$$\bar{p}_\infty(S') = (1-k')\,\frac{\pi\theta_1}{\pi\theta_1+\bar{\pi}\theta_2} = \frac{N_{(I)}\left(\dfrac{\pi\theta_1}{\pi\theta_1+\bar{\pi}\theta_2}\right)}{N_{(I)}+N_{(S'-I)}}, \quad (10)$$

where $N_{(S'-I)}$ is the number of elements in $S'-I$ and $N_{(I)}$ is the number of elements in I, one can empirically alter the number of elements in $S-I$ and $S'-I$, and, by evaluating empirical asymptotes, determine the extent to which uncontrolled stimulation is weighted within I. Such a specification makes contact with information theory, since intercept elements constitute "noise" in discrimination.

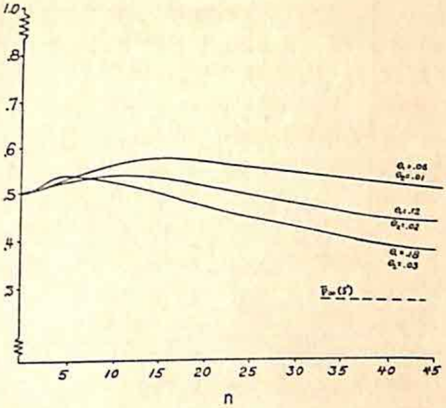


FIG. 2. Effects in appearance and magnitude of the S' peak as a function of the value of the slope constants.

The rapidity with which the increase in $\bar{p}_n(S')$ occurs, when it does occur, depends upon the slope parameters $(1-\theta_2)$ and $(1-\pi\theta_1-\bar{\pi}\theta_2)$. As shown in Fig. 2, the smaller the θ values, the more delayed will be the appearance of the peak within the series, and the greater will be the rise in the peak.

One final comment is perhaps due. The term *stimulus induction* has been used synonymously with *stimulus generalization* (**9**). It seems reasonable to propose that induction be reserved in the present analysis to refer to the dynamics of successive conditioning and extinction within the intercept. This leaves generalization to the more static conditions treated in the Bush-Mosteller paper. Such a definition seems more in line with Skinner's original usage of the term.

DISCUSSION

Since θ values depend both upon organismic and environmental factors, manipulation of either of these has corresponding effects. Specifically, if the sampling ratio is reduced, either by increasing the effective stimulus field or by interfering surgically with the organism's discriminatory apparatus, the S' peak will be delayed in its appearance within the stimulus

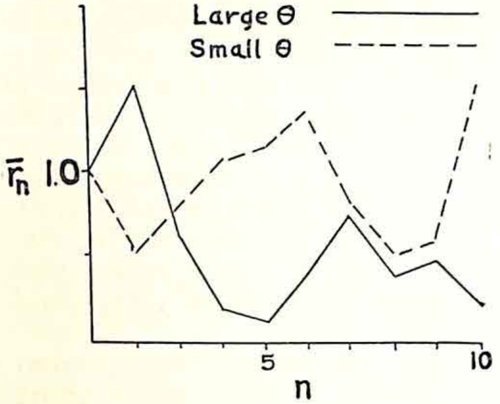


FIG. 3. Displacement of empirical S' peaks resulting from reducing the total number of stimulus elements to be discriminated.

series. On the organismic side, Dr. Karl H. Pribram, after inducing temporal lesions in rhesus monkeys, reports both the expected delay in the appearance of the S' peak and the expected differential in magnitude of the peak in the behavior of operated animals as compared with normals.[3] Data from previous studies (**7**, **8**) support predictions of the model concerning the appearance of the peak as a function of manipulations of the stimulus field. Figure 3 shows the effect upon the position of the peak of increasing the sampling ratios. The increased sampling ratio was produced by reducing the total number of elements to be sampled in S'; the effect is an earlier appearance of the peak.

An experiment was performed explicitly to evaluate the reality of the model's predictions with regard to presence and absence of the peak.[4] Ten subjects were run in each of two groups. In each group an equal number of cards with S and S' patterns were shown the subjects. The subjects were required to state when they believed an S pattern was being presented; they were informed when they were correct. The series of patterns was shuffled between experimental runs, so that each subject was exposed to a different series of patterns. The patterns were composed upon cards upon which were drawn ten circles in two lines of five circles each. In Group I, one particular circle was always filled in if the pattern was an S. A different circle was filled in to define S' instances. Three other circles, comprising the intercept, were filled in in various combinations of one, two, or all three; and these were presented equally often with either S or S' cards. In Group II, the S and intercept circles were filled in, as in Group I, but S' of this latter group consisted of the remaining six circles, one of which was always filled in. In brief, the technique is the same as has been used in a free operant situation.[5]

Under both conditions, the subject is faced with a guessing problem at the outset, hence an estimate of .5 for $\bar{p}_0$ seems not unreasonable. The model makes explicit determinations of several constants for these experimental conditions, as is shown in Table 1.

Consequently it remained to estimate but a single parameter; either θ_1 or θ_2 in the Group II condition, since both θ_1 and θ_2 of Group I are equal to

TABLE 1

DETERMINATIONS OF THEORETICAL CONSTANTS

Group	Constant					
	θ_1	θ_2	$\bar{p}_0$	k	k'	$\frac{\pi\theta_1}{\pi\theta_1 + \bar{\pi}\theta_2}$
I	.18	.18	.50	.25	.25	.50
II	.18	.03 *	.50	.25	.67	.86

* Estimated empirically.

---

[3] As reported in his unpublished manuscript entitled, "On the neurology of thinking."

[4] Acknowledgment is due E. P. Stichman who conducted experimental work in the early phases of this study.

[5] A fuller report of this experiment is being prepared by the writer.

θ_1 of Group II; and since all of these are six times as great as θ_2 of Group II, θ_2 of Group II was estimated to be .03. Two principal points were examined in this experiment: one being the prediction that the equal θ case should generate S and S' curves which were symmetrical about $\bar{p}_0$; S and S' curves in the unequal θ case should, however, be characterized by a more rapid acceleration to S asymptote than is true in the equal θ case, and an S' peak should occur which should be followed by a slow decrease to S' asymptote. Figure 4 demonstrates the gross relation obtained between theoretical curves and the data gathered in this experiment.

The second prediction was that, aside from the gross characteristics of the curves, reasonably accurate predictions concerning exact response probabilities should be rigorously made by the model. The theoretical distribution for the curves is not known, so a χ^2 test of goodness of fit was performed. In Group I the levels of confidence for S and S' were $(.6 < p < .7)$ and $(.7 < p < .8)$, respectively. In Group II they were $(.9 < p < 1.00)$ for both curves. Admittedly the fit is best for the condition from which the estimate of θ_2 was taken, but fitting all four curves through the estimation of one parameter with the degree of precision that was attained in this experiment is encouraging. Certainly the model is distinguished from others in the uniqueness of its predictions concerning a given experiment.

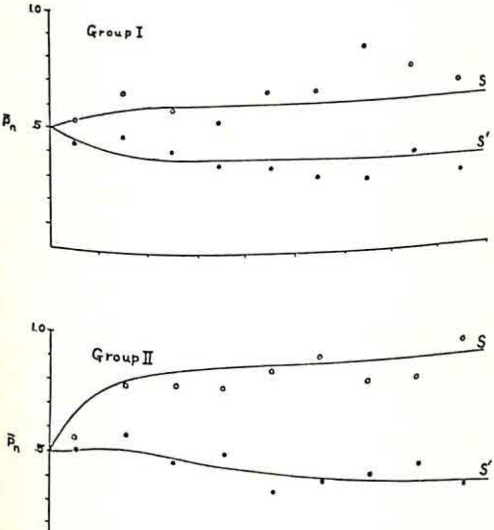


FIG. 4. Theoretical curves and experimental results where $\theta_1 = \theta_2$, Group I, and where $\theta_1 = 6\theta_2$, Group II.

In examining what happens in the experimental setting, one observes activity reminiscent of Kantor's analysis of a response within a behavior segment (**10**). Kantor has defined a response as being composed of three reaction systems: the attending, the perceiving, and the consummatory. Explicitly, in this and similar experiments, the subject is required to orient toward, or attend to, the stimulus patterns, to respond visually, and to emit some operant response. In the model, both conditioning and extinction are regarded as probability transfers between an observed class of behavior and some mutually exclusive classes of behavior, without examining the nature of the non-observed classes. As conditioning takes place, competing behavior decreases in strength. As extinction occurs, competing behavior takes the place of the experimentally measured class. The kernel of the argument is that during extinction, as competing behavior increases in probability, the competing behavior carries with it its own postural or orientational aspects. These changes in orientation, or physical position of the subject, systematically increase the relative size of the $S'-I$ set, thereby reducing the probability of sampling an S' element. This has the effect of distorting estimates of both the sampling ratio and the constants of proportionality.

It logically follows that a corresponding distortion should occur in S or conditioning trials. The value of θ should increase as conditioning progresses. Estes has already indicated that this is the case (**4**). Hence the more rapidly conditioning progresses, the more rapidly will it progress. This may have some relevance to analyses of "learning sets." At the same time, if extinction is rapid at first, it may be expected to interfere with itself, so to speak, since sampling efficiency is reduced.

The second distortion so introduced is in the constants of proportionality, k and k'. Initially, uncontrolled stimuli are present as a part of the intercept. As the systematic changes take place, as described above, which translate much of this uncontrolled material into $S'-I$, I is reduced relative to both S and S'. It is here that some behavioral meaning can be given the Bush-Mosteller statement about the diminishing intercept. Such a displacement tends to exaggerate the departure of the S and S' curves from $\bar{p}_0$; thus an analysis is possible which accounts for the occasional instance of near-perfect discrimination in a manner which is preferable to some others which have been invoked to account for this phenomenon elsewhere. The discrimination operator D in the Bush-Mosteller model is clearly a mathematical device introduced with little rationale other than the fact that it produces curves which look more like discrimination curves. Restle's adaptation model is also designed to handle this problem, but it unfortunately introduces more elaborate assumptions concerning the conditioning process (**12**, **13**). The present analysis would seem preferable since it involves no more complex assumptions concerning the conditioning process than does the original Estes-Burke theory, and since exact experimental tests of its correctness are immediately evident.

If orientational changes do in fact distort the parameters k and k' as a function of trials, they should do so in a given experimental setting to a greater or lesser degree, depending upon the number of experimental elements actually manipulated within the three subsets. That is, the fewer the number of experimental elements, the greater is the relative contribution of the uncontrolled stimulation and the greater the distortion produced in the constants of proportionality when the uncontrolled elements shift. The more elements existing within the sets, the less is the relative contribution of uncontrolled stimulation, and the less will a shift distort k and k'. In the experiment presented above, the greatest distortion does indeed exist where the fewest elements are involved. While this is only suggestive that the present analysis is correct, current experimental work is under way to evaluate this phase of the problem in detail.

One technique to reduce the systematic distortion has been proposed by Estes, where intertrial intervals are increased. One might reason that this provides ample opportunity for the subject to sample everything in the environment, such that any systematic sampling in S' would exert little relative effect. In regarding orientational changes as stimulus producing, this reduces to the analysis above in that the total number of stimuli is increased. Certainly a mapping of the response space is a sorely needed development in further enterprises of this kind.

The alternative to this analysis would seem to involve the introduction of additional complicating assumptions; while any number of mathematical devices can readily be assumed which more or less solve the curve-fitting problem, the experimental approach remains the more satisfying.

SUMMARY

A particular examination of the S' peak from the standpoint of an extension of the Estes-Burke statistical learning theory supposes the peak to be generated by simultaneous additive and subtractive processes working upon response probabilities associated with stimulus elements within and outside the intercept of the S and S' subsets. Certain exact experimental predictions are made and examined, and some questions concerning its use in developing a program of research are raised as it compares with other current approaches. In the light of available evidence, extensions are suggested for broadening the scope of the present analysis.

REFERENCES

1. BURKE, C. J., ESTES, W. K., & HELLYER, S. Rate of verbal conditioning in relation to stimulus variability. *J. exp. Psychol.*, 1954, **48**, 153–161.
2. BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
3. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. ESTES, W. K. Theory of elementary predictive behavior: an exercise in the behavioral interpretation of a mathematical model. *Mathematical models of human behavior—proceedings of a symposium.* Stamford, Conn.: Dunlap and Associates, Inc., 1955.
5. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
6. ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
7. GREEN, E. J. Concept formation: a problem in human operant conditioning. *J. exp. Psychol.*, 1955, **49**, 175–180.
8. GREEN, E. J. Stimulus-variability and operant discrimination in human subjects. *Amer. J. Psychol.*, 1956, **69**, 269–273.
9. HILGARD, E. R. *Theories of learning.* 2nd ed. New York: Appleton-Century-Crofts, 1956.
10. KANTOR, J. R. *Principles of psychology.* New York: Knopf, 1926. 2 vols.
11. KEMENY, J. G., SNELL, J. L., & THOMPSON, G. L. *Introduction to finite mathematics.* Englewood Cliffs, N. J.: Prentice-Hall, 1957.
12. RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.
13. RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.
14. SKINNER, B. F. *The behavior of organisms: an experimental analysis.* New York: D. Appleton-Century, 1938.

(Received June 28, 1957)

ANNOUNCEMENT

JOURNAL OF EDUCATIONAL PSYCHOLOGY

This journal will now be published by the American Psychological Association. In 1958 it will become a bimonthly; issues will appear in February, April, June, August, October, and December. Contents include articles on problems of teaching, learning, and the measurement of psychological development.

All back issues and subscriptions up to and including the May 1957 issue are the property of Warwick and York, Inc., 10 East Centre Street, Baltimore 2, Maryland.

Subscription $8.00
(Foreign, $8.50)

Single
Copies, $1.50

Direct new subscriptions and renewals to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

CONVIVIUM
ESTUDIOS FILOSOFICOS
UNIVERSIDAD DE BARCELONA

CONVIVIUM appears twice a year, and contains articles, notes and discussions, book criticism, and an index of periodicals. Articles are printed in the Latin languages, with a summary in English and German.

Editor: Jaime Bofill Bofill (Professor of Metaphysics)

Price: Single copy, 60 pesetas (Foreign, $2.40)
Subscription, 100 pesetas (Foreign, $4.00)

Mailing address: Sr. Secretario de CONVIVIUM,
ESTUDIOS FILOSOFICOS
Universidad de Barcelona
Barcelona, Spain

VOL. 65, NO. 2 MARCH, 1958

BUREAU OF EDNL. & PSYL. RESEARCH LIBRARY

THE PSYCHOLOGICAL REVIEW

SENSATION AND PERCEPTION IN AN OBJECTIVE PSYCHOLOGY[1]

C. H. GRAHAM

Columbia University

This paper will take up what might have been called, in an earlier day, a systematic treatment of the topics of *sensation* and *perception*. In particular, it will consider their places within the formal structure of an objective or behavioristic psychology.

The development of such a psychology has been a major historical occurrence of the last 35 years, and, in some respects, it has achieved a considerable level of maturity. Despite the fact that behavioristic descriptions of perception were advanced at a relatively early date (see, for example, 9) the area of perception still remains a difficult formal problem for systematization within the framework of an objective psychology. The reason for the refractory nature of the subject probably lies in the complexity of its relations as contrasted with the deceptive familiarity of its terminology. In any case the general topic requires analysis. Specifically this paper will discuss the following five topics: types of stimulus-response relations encountered in perception, the problem of the independent statuses of sensation and perception, the problem of descriptive and explanatory terminology, some recurring considerations in perception, and perception and psychological theory.

It is hoped that the discussion of these seemingly miscellaneous topics may take on the character of an integrated account. In all of this we shall not deal with explanatory principles as they relate to specific processes. We know that the area of perception is rich in such principles, particularly as respects vision and audition. What will be done is to consider some general characteristics of the large over-all area and its relation to other fields of psychology. The attempt will be supplemented by brief references to a symbolic notation, in terms of stimulus and response, that has been discussed previously (3, 4, 5, 6).

TYPES OF STIMULUS-RESPONSE RELATIONS IN PERCEPTION

In analyzing the sort of behavior that can be encountered in experiments on sensation and perception, let us start with a particular type of experimental situation, one that applies to the experiment dealing with introspective reports of sensations or perceptions evoked by

---

[1] The preparation of this manuscript was supported by a contract between the Office of Naval Research and Columbia University. Reproduction in whole or in part is permitted for any purpose of the United States Government. The account encompasses, with minor modifications, the content of the writer's presidential address to the Eastern Psychological Association at Atlantic City on March 23, 1956.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated

particular objects or events involving the objects. Concretely, the question arises: What do I have to do in this situation to get a subject to give me a phenomenal or introspective report in the presence of certain objects that I supply?

Proper procedure in such a situation demands, of course, that I instruct the subject to give me a phenomenal or introspective report and that I supply the appropriate objects in their appropriate relations. The consequent behavior on the part of the subject will be the emission of a sequence of words—words that are, we presume, functions of the subject's past history of reinforcements (by parents and other persons) in a sequence of discriminative trainings as discussed, for example, by Skinner (**15**). These verbal responses are often said to be "descriptive" of stimuli. In situations such as the one described, the experimenter hears such words given by the subject as *moving, color, short, white, object, large, red, duration, size, movement, intensity, texture,* and a whole gamut of expressions that "apply" to the character of the stimuli. Since these words were conditioned in response to aspects of stimuli that are now present, they are "appropriate" responses, both to the subject and the similarly conditioned experimenter. In all of this, the place of the instruction stimuli in restricting the possible number of responses remains to be understood. To say that they provide the necessary "sets" is not very informative.

It would be of considerable interest to compare a subject's behavior under instructions to introspect or give a phenomenal description with behavior under instructions to describe objects. Certainly the subject may manifest different behaviors as exemplified in "the subjective attitude" and the "objective attitude." To what extent differences are apparent in his behavior is a question on which not very many relevant data exist at present. While the problem should not deter us, it may have later implications for a behavioristic account of perception.

In experiments of the sort under consideration—that is, experiments on phenomenal appearances or introspective reports—it is an empirical fact that the instructions used do not greatly restrict responses. The latter may exhibit a great many degrees of freedom—that is, many different words may be spoken. In addition, neither stimuli nor responses are usually measured in the observations on phenomenal appearances or introspection.

The problem of analyzing and classifying the responses in an experiment of this type is, of course, one of the major problems of psychology. In particular, the problem of analysis and classification of responses does not involve accepting them as understandable conversation, but is one of formulating their uniformities and rules in a system where they are taken *as* behavior. This sort of analysis provides the anomaly that conversation is not viewed as such, but is considered in terms of other words or symbols—that is, scientific description that formulates the rules of the subject's conversation. We deal here, of course, with the problem, discussed by Bergmann and Spence (**1**), of the subject's language and the experimenter's metalanguage.

Let us now turn to another type of experiment, one that, as contrasted with the experiment on introspection, shows measurement of both stimuli and responses as well as considerable restriction of both. Of course, this type of experiment is scientifically at a somewhat higher level of development than the introspective experiment, due to the fact that in the past its permissible stimuli and responses have been isolated and measured in one way or another. In this situation the type of relevant procedural question is: What do I have to do to get a subject to say, with a certain degree of probability, "I see a light," or its converse, "I don't see a light"? The subject, in the course of any single stimulus presentation, can only give one of the two possible responses (or under appropriate conditions, their "equivalents," *Yes* or *No*).

This type of experiment is the classical psychophysical experiment. It turns out, in this situation, that light intensity must be manipulated, and we can plot some measure of response—for example, its probability of occurrence—against some appropriate function of intensity. This procedure has been followed in Fig. 1.

This figure gives the psychophysical data of three subjects for an experiment on quantum requirements at threshold by Hecht, Shlaer, and Pirenne (7). Such a function is a stimulus-response relation that shows how a measure of response varies with a controlling stimulus variable. From such a relation it is possible to obtain a new datum, a value of the stimulus variable that corresponds to a given probability of response occurrence—for example, the 50% value. This new and extremely important datum is familiarly known to us as a threshold.

The two categories of the experiment that have been discussed, the phenomenal report and the psychophysical, lie at the extremes of an array of psychological experimentation, the array being ordered on the basis of degree of stimulus and response restriction. The introspective experiment involves relatively unrestricted stimuli and responses, while the psychophysical experiment usually involves a high degree of restriction in both.

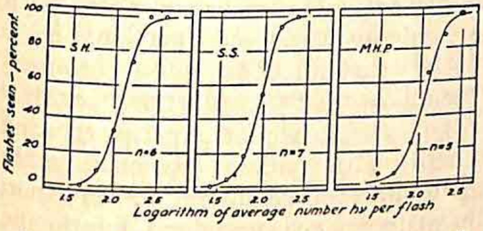


FIG. 1. Frequency of seeing functions obtained on three subjects by Hecht, Shlaer, and Pirenne (7) in their experiment on quantum requirements for visual threshold.

Other types of experiments take up positions between the extremes of such a continuum. One may have in mind particularly the type of experiment involving absolute judgments, estimates, or naming behavior. This type of experiment deals with the relation between environmental stimulus variables, often measurable in some theoretical numbering system taken, for example, from physics, and large families of variable responses. This is the sort of experiment that involves such procedural questions as: What do I have to do to make a subject say "I see five dots," or "This is red," or "This has a value of three on a scale of one to five"? The response given in this experiment is one that has developed during the subject's life history in the presence of objects. "Naming a color" and "giving a numerical estimate of brightness" are examples of such responses.

Figure *2* gives a set of hypothetical curves for color naming in the spectrum. Each curve gives, as a function of wave length, the frequency of occurrence of the response that applies to the curve.[2]

As contrasted with the method of constant stimuli with two categories, where only two responses can be elicited

---

[2] The curves of Fig. 2 were not drawn on the basis of data. They were constructed on the basis of "reasonable guesses" relating to other kinds of color discrimination. Since this paper was written, Aleeza C. Beare, working at Columbia, has made experimental determinations of color naming functions at two levels of intensity. The functions are, in fact, quite similar to the hypothetical ones shown in Fig. 2.

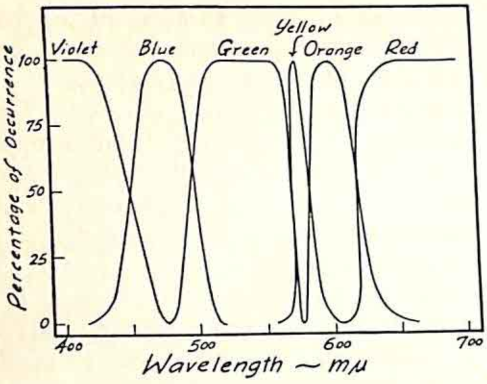


FIG. 2. Hypothetical curves for frequency of occurrence of color names in various parts of the spectrum.

no matter what the value of the discriminative stimulus, the method of absolute judgment allows for the production of a great number of responses. For example, in the naming of a color, a considerable change in wave length can result in many changes in response. Many problems beset our understanding of this sort of behavior function. In particular, we cannot specify, as yet, quantitative continua that represent relevant *amounts* of differentiated responses.

A number of undesirable ways of talking about the results of the experiment on absolute judgments hinge on this problem. Here one may mention two examples. First, there is the device, harmless enough if we keep in mind what is being done, that has sometimes been used to imply response continuity. This is the device of assuming a response continuum as the correlate of physical dimensions of the stimulus, as when we say that "orange" (rather than the wave lengths that evoke the response "orange") is next to "red" in the spectrum (rather than the wave lengths that evoke the response "red"). Secondly, there is the procedure of treating numerical estimates as if they were numerical data. For example, the word *ten* emitted by a subject in an experiment on estimated brightness may be treated, erroneously, I think, as if it were the quantified outcome of a measuring operation; the symbol for response (accepted by the experimenter as a number) is added to the list of other numbers of the same sort, and the experimenter says that an "average" response is finally computed. No independent operations of measurement have been applied under these circumstances to the subject's responses.

In any case, it should be pointed out that we can theoretically derive critical values—threshold values, if you will—from certain types of experiments on absolute judgments. For example, we can find, on the average, how many dots are required for the response *five*. We can also find out how many are required for the responses *four, three, two, one,* etc. Similarly, in the case of estimates of brightness, one can find the value of luminance that is on the average correlated with the response *ten*, for example. We do not find what the average response is for the same luminance.

The concept of threshold, i.e., a critical value of stimulus corresponding to a given response criterion, to which previous reference has been made, need not be elaborated here; it is an old and pervasive principle of psychology. Nevertheless, it will not be out of place to point out that, by virtue of this quantity, it is possible to treat as a single datum the outcome embodied in a total psychophysical function, and so derive what might be called perceptual functions—that is, functions obtained from a sequence of experiments concerned with finding thresholds under many different conditions. Such experiments pose such procedural questions as the following: What do I have to do to get a subject to say "This light is brighter than that one" under conditions where I systematically vary the standard light?

Figure 3 shows such a sequence of psychophysical curves obtained by Mueller (14) in his determination of the just discriminable increment in intensity at various intensity levels. Each curve gives a threshold Δ*I*, the critical value of Δ*I* taken as corresponding to the 50% response occurrence.

In a further extension of the data, we determine how the second-order datum, the threshold Δ*I*, varies with changes in the standard light. The function obtained in these circumstances shows how determined values of Δ*I* (each one based on a psychophysical function) vary with the intensity of the standard light.

Figure 4 represents such a function. It is the curve for Mueller's threshold values of Δ*I* as a function of the background intensity to which Δ*I* is added. Such a function, which we may term a perceptual function, shows how a critical value of stimulus, the threshold, varies as a function of a controlling variable.

It is of interest to examine the present analysis in terms of a symbolic formulation that has been used (Graham,

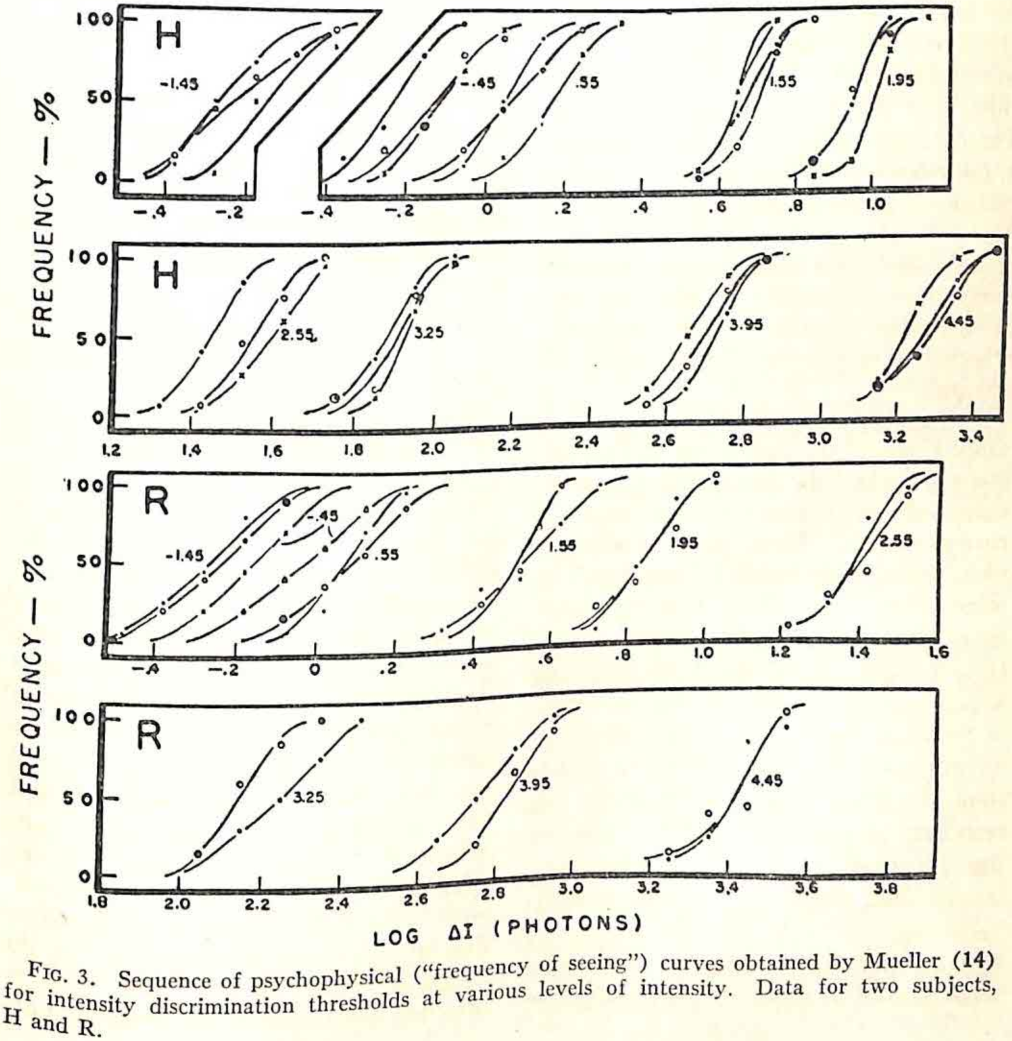


FIG. 3. Sequence of psychophysical ("frequency of seeing") curves obtained by Mueller (14) for intensity discrimination thresholds at various levels of intensity. Data for two subjects, H and R.

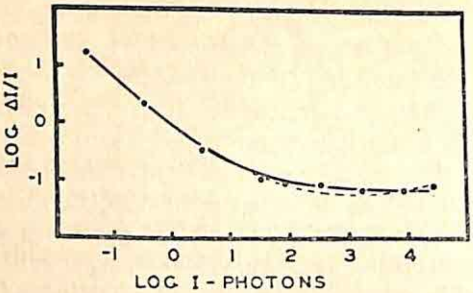


FIG. 4. Intensity discrimination ratios ($\Delta I/I$) at various levels of intensity based on the critical ΔI's obtained by Mueller (14) at the 50% frequency of occurrence values in the curves of Fig. 3.

3, 4, 5, 6) to specify some classifications of relations among stimulus, response, and conditions of the subject. Specifically, the variables of the present analysis must be examined to see if they can be described as special cases of the general relation

$$R = f\,(a, b, c \ldots n \ldots t \ldots x, y, z) \qquad [1]$$

> This relation expresses the assumption of a behavioristic program: response is a function of certain specifiable variables. In particular, the first letters of the alphabet (a, b, $c \ldots$) refer to properly specified aspects of stimuli; the last letters ($\ldots x$, y, z), to properly specified condition of the subject (physiological and inferred, including the effects of instruction stimuli); R, to response; n, to number of presentations; and t, to time. The terms are not always independent of each other (Graham, **6**, p. 62).

It is important to realize that x, y, and z may be, among other things, the consequences of past stimulation.

The stimulus-response data of a psychophysical experiment can be represented, at least for the cases of the methods of constant stimuli and limits, by the equation:

$$R = F(a, x_1) \qquad [2]$$

This equation is a special case of Equation 1: response is a function of an aspect a of the stimulus, with x_1, the effect of instructions, explicitly specified as constant. The same sort of analysis can be applied to the data of absolute judgments. Here it must be admitted, parenthetically, that to a considerable extent the discussion here presented depends for its validity on the choice of the methods of constant stimuli and limits, both with extensive series of observations. When the method of adjustment is used, response measurement becomes difficult.

A combination of psychophysical experiments provides the data for a perceptual function as in

$$a_c = \Phi(b, x_1, R_1) \qquad [3]$$

where a critical aspect of stimulus a_c is represented as a function of another aspect of stimulus b. In this case x_1, the effect of instruction, and R_1, the criterion response, are maintained constant. In other experiments, a could be determined as a function of such variables as n, t, or y.

Equations 1, 2, and 3, and similar ones give a convenient classificatory account of the types of relations found in the psychophysical experiment. What about the relations of the experiment on phenomenal appearances? Is there a distinction between the data of the introspective and psychophysical experiments? Certainly there seems to be in two respects. First, both the stimuli and responses of the introspective experiment are relatively unrestricted. They depend for their occurrences upon the behavior of a subject who is responding and being stimulated in many different ways. Secondly, neither stimuli nor responses are measured in the introspective experiment, and so a quantitative description of stimulus-response dependence is impossible. While it is true that stimuli and responses are presumably related from a rigorous mathematical point of view, the statement has little utility in the absence of knowledge of the specific relation. Nevertheless we can use the reports of such an experiment at a conversational level to provide us with hypotheses to be tested in the more formalized psychophysical experiment. We may also examine the reports themselves with a view to analyzing them in the context of a descriptive or explanatory system of language.

It is to be hoped that we shall not long remain ignorant of what the subject is doing in the introspective experiment, and, in fact, in any circumstance in which he talks.

THE SEPARATENESS OF SENSATION AND PERCEPTION

Up to the present time nothing has been said about differences that may exist between sensation and perception. For all practical purposes the two topics have been considered to be one. The question now arises: Can they, in fact, be differentiated on the basis of appropriately selected criteria?

Sensation and perception were accepted as different in the general framework of a structural psychology. Sensations were taken to be the irreducible elements of consciousness; and perceptions, the carriers of meaning, were judged to be aggregates of sensations and images (Titchener, 16). Exponents of phenomenological description took another tack. They regarded their particular interest as concerned with the phenomenal world or the phenomenal field (cf., e.g., Gibson, 2; Katz, 10; MacLeod, 13), depending on the different attitudes taken in observation. Of course, Gestalt psychologists have, from their beginnings, disclaimed the concept of sensation *in toto;* they deal with stimulations and their consequent patterning (cf. Koffka, 11; Köhler, 12).

Despite the great differences exhibited by these various psychologies in the details of their treatments of conscious experience, they do demonstrate, at least superficially, an emphasis on the importance of the patterning of units in perception—units that are variously conceived to be sensations, stimulations, or cues.

In addition to their usage within the context of psychologies of experience, the terms *sensation* and *perception* are sometimes used in a different, possibly unwarranted manner. In consequence of their preoccupations with the nature of stimulation, psychologists often now use the term *sensation* to include whatever physiological theory and data of sensory processes may seem to be applicable in a given circumstance. Sensation has come to be thought of as essentially physiological, whereas perception, which is relatively poor in physiological theory, is considered to be psychological.

In view of the long historical background of the terms *sensation* and *perception,* one may legitimately ask: Are either or both terms needed in an account of behavior? It must be admitted that one could use many terms as convenient classificatory devices in discussing behavior. For example, the term *discrimination* might be used to cover the class of functions that are of interest here. It is probable, however, that the latter term would usually be interpreted in a more restrictive sense than would a term that covers the array of functions here discussed, characterized as they are by the fact that response is treated as a function of stimulus, with conditions of the subject due to past reinforcements at constant or limiting values. (Of course this statement does not mean that we are not interested in past history parameters! In the latter circumstances, response is treated as a function of conditions of the subject with stimulus constant. But of course the latter descriptions have their own designation: learning or conditioning functions.)

In any case, at least for the present, the term *perception* will be retained to refer to classes of stimulus-response functions that represent constant values of reinforcement parameters. Does a basis exist for retaining the term *sensation* independently of *perception?*

Let us examine the problem. The behavior observed in all perception experiments (introspective, phenomenological, or psychophysical) involves stimulus discriminations and response differentiations. The correlated terms in the observed stimulus-response functions are the experimental data. In experiments that provide such data, an instruction stimulus is given, ideally only once; thereafter, a stimulus aspect is changed and certain responses are observed. Variations in a controlling parameter of the stimulating circumstances give rise to correlated response occurrences or changes. This context seems to cover all of the functions classifiable under the heading of perception, with no surplus functions requiring a new heading. Thus, on the side of description of experimental operations, it seems that no additional term is needed. It is true, of course, that the needs of theory cannot be finally prejudged. It might be that some theoretical analysis could provide appropriate criteria for the use of two or, in fact, more terms to specify processes of different orders of complexity. So far the use of two terms does not seem to be based on mutually exclusive criteria. We seem to do well enough with one.

THE PROBLEM OF TERMINOLOGY

The analysis of experiments in perception and their related concepts pose an important practical and theoretical question: Shall we dispense with many of the terms of an earlier vocabulary and try to develop new terms for a behavior science, or shall we define new meanings for the old terms? It may be too early to tell whether an account of behavior can more advantageously use old terms, appropriately respecified, or new terms, divested of certain historical contexts. In any case, it will probably do no harm to see what considerations might arise from the use of one set—specifically, the old terms. Such an examination does not now imply an ultimate preference for one system as opposed to the other. It may, however, be supposed that a sequence of such examinations would at a later time provide us with a basis of choice.

Let me take some examples from the field of color. In this area it might conceivably be possible to get along without such words as *color, brightness, saturation,* or *hue,* provided that we could specify the end terms of a given discrimination. Experience, however, shows that, even if we could always (as we cannot) express uniquely the energy relations that correlate with a given response (and vice versa), the use of such terms would be cumbersome. For this reason we shall here consider the roles of such words as *hue, saturation, brightness, color* and *chromaticness*. We shall not argue the question of whether these names are labels, intervening variables, or hypothetical constructs. The applicable terms will probably be chosen on the basis of the theorist's attitude.

Consider, first, the use of the term *hue*. This term is to be understood as either a label for or as an inferred effect (behavioral, physiological, or mathematico-representational) in the following stimulus-response sequence: (*a*) instructions to a subject who has had a past history with the vocabulary represented in the instructions, (*b*) the presentation of radiant energy to the subject, and (*c*) the subject's responses. It turns out, as a matter of empirical fact, that wave length is the most important variable in (*b*) for hue discrimination.

The dependence of hue discrimination on wave length may have been established early in such observations as the ordering of stimuli by a person when he is instructed to arrange colors on the basis of hue. The subject is said to discriminate differences in hue when he gives one response (for example, a hue name) to a radiant flux of one narrow wave-length band and another response to flux of a different wave length. Another type of correlated observation involves the subject's giving (*a*) one response ("No, there is no difference in hue") to a small difference between wave lengths of a pair of stimuli, and (*b*) another response ("Yes, there is a difference in hue") to a larger difference. The threshold wave-length difference is evaluated at the appropriate frequency of occurrence of the two responses.

It will be observed that the word *hue* comes into play at least twice in the sequence *instructions, stimulus, response:* It occurs as part of the instructions (as in the statement "arrange these colors on the basis of hue"); and it is the term for the inferred effect relating stimulus and response, or the label for the relation.

The word *hue* used in the instructions (and reacted to by the subject) is analytically a different word from the word applied to the discrimination. The difference in the two words is comparable to the difference discussed by Bergmann and Spence (1) when they contrast the "social" language of the subject and the experimenter with the metalanguage of the experimenter as a scientist. Possibly, the difference could be specified by the use of subscripts, as in hue_i and hue_d, where the former word represents the term in the instructions, and the latter the term applied to the discrimination. The former term controls a subject's activity; it "tells him what to do"; it may imply little or no theoretical context. The latter term involves whatever meanings may be attached to it by its ramifying empirical and theoretical connections. Thus, for example, it may imply an elaborate context of physiological mechanisms.

Considerations comparable to those holding for the word *hue* exist with respect to the word *saturation.* The term represents the inferred effect or relational label holding between such a stimulus variable as colorimetric purity and differential responses that exist when a subject is instructed to order colors on the basis of, for example, "paleness of color."

Differences in brightness may be discriminated by processes analogous to those holding for hue discrimination. The subject, under instructions to arrange stimuli according to differences in *brightness* or to indicate *brightness differences,* gives different responses (for example, the words *dim* and *bright*) to different luminances, or he may differentially signal a difference between luminances. Such discriminations are said to be *brightness discriminations.* The word *brightness* appears in the double sense characteristic of the word *hue.*

The word *color* represents a generic type of discrimination based on a combination of wave length, colorimetric purity, and luminance. The subject, properly instructed, can respond by naming different colors or with such differential responses as "Same color" or "Different."

The term *chromaticness* is of considerable interest, based as it is on a very definite type of representational theory: the chromaticity diagram. Chromaticness theoretically represents a generic type of discrimination based on a combination of wave length and a quantity called excitation purity. It thus may be taken to be a combination of hue and saturation, with brightness removed from consideration. As a concept it is based on a highly abstract mathematical representation, and its correlated responses have not as yet been carefully considered.

The word *perception* itself requires some discussion. The word does not in itself refer to an inferred effect having the significance of an intervening variable (as has been the case with the words previously discussed). It is a class name for all such variables. Presumably, it is to be understood as the class of all relations conventionally subsumed under the word *perception*. (We shall not here go into the problem of formal bases for this subsumption.) In any case, we should not expect to hypothesize about a generic *process* of perception as we might about a specific component process.

This discussion of representative terms in the field of color, some being taken directly from an older psychology, will probably show that the latter terms can be used in the vocabulary of a behavior science. Whether they will serve there more suitably than new terms is an open question. Would it be better, for example, to speak of wave-length discrimination rather than hue? On this problem at least one thing may be said: Comparable aspects of wave-length discrimination show different dependencies on wave length as functions of other variables. Put in another way, hue varies with other factors than wave length. The difficulty posed by this fact can be surmounted, but possibly only at the cost of some terminological cumbersomeness.

SOME RECURRING CONSIDERATIONS

Let me now take up some recurring considerations in the study of perception. In particular, I feel that the old empiricistic-nativistic controversy may still have some implications for present-day studies. I am not pleading for a renewal of this ancient argument, but do desire a clarification of some of its concepts in the interest of benefiting our present understanding.

Any situation that involves a discrimination, an estimate, a denotation, or a naming response employs a subject with a system of responses. Many, or most, of the responses are the end-products of the subject's history of acting in a complex and changing environment. Now it is never shown in researches on the intact organism that the subject's past history is irrelevant. What may be shown is the fact that, in certain situations, past history variables do not establish the *form* of a given function. In a word, such variables exist at constant values. Under other circumstances, they do influence the form of a function; they are contributing variables. The identification of both classes of function is a matter for experiment.

Past history variables are important in some types of perceptual behavior, and they are now receiving increased attention. Studies of motivational factors in perception are being conducted at an accelerated pace. Despite some criticisms that can be made of certain recent experiments, nevertheless, in general they support a permissible modern interpretation of Helmholtz's position (8) on unconscious inference: discriminative behavior is influenced by past conditions of the subject. The study of the latter effects is in one of the two major directions of perceptual investigation. In pursuing this line, answers may be sought to such questions as the following: What do genetic studies tell us about the development of certain discriminations? What sorts of training procedure can be shown to influence discriminations and how do they do it? How will different instructions change the responses of the subject? What are appropriate dimensions for the quantification of instructions?

The other major rank of perceptual study is, of course, an old one. It is concerned with the problem of how systematic variations in stimulus conditions influence discriminations. This area has had its problems and controversies, but we need not revive them now. It is sufficient to say that the necessity for advancing our knowledge of the energy bases of discrimination to more refined levels remains with us. It is true that some quantitative experiments have been done in the past, but it is also true that experimental parameters have not always been varied over a large range and in a systematic manner. Put briefly, many perceptual experiments of the past have often been demonstrational rather than analytic, particularly in cases where variables other than intensity and wave length have been the main centers of interest. One thing the study of perception needs is adequate data to provide a firm base for theory. In the future it may probably turn out that many interesting demonstrations may be seen for what they are—special cases of much more general quantitative principles.

PERCEPTION AND PSYCHOLOGICAL THEORY

It seems interesting and important that we can go a considerable way in classifying some of the discriminative functions in perceptual behavior and that we may designate two directions which appropriate studies may take: investigating (*a*) conditions of the subject, and (*b*) stimulus conditions.

It is probably true that the studies of conditions of the subject (in the first direction) must be brought into relation to learning and conditioning. To put it briefly, such studies must make meaningful contributions to our understanding of behavior modification, the area where past history variables are of primary concern.

Studies of the stimulus control of behavior (in the second direction), when properly made, usually do not show the influence of past history as a contributing variable. The experimentalist and theorist in this area has the responsibility of relating his findings to an already existing precise body of formalized information. In a sense, then, the experimentalist in this area must do more than classify the behavior functions. He should be sensitive to the need for relating his findings to appropriate theory, sometimes in the vocabulary of psychology, sometimes in the vocabulary of physiology, and often in the terminology of mathematics. By his contributions the psychologist interested in the stimulus control of behavior can provide sounder bases for the interpretation of the often admirably conceived but disconnected hypotheses that now exist, and, in a larger context, advance the synthesis of these hypotheses into a more inclusive theory.

Finally, it is clear that an ultimate synthesis of information obtained in future investigations—i.e., from studies of (*a*) conditions of the subject and (*b*) the stimulus control of behavior—will be required for purposes of a more embracing theory. Data from the first direction will tell us how discriminative behavior develops and what are the conditions for its development; and data from the second direction will tell us how such discriminations are controlled and limited. Such a synthesis will make the study of perception coextensive with the general study of behavior.

SUMMARY

This paper discusses five topics that are of systematic importance in the treatment of sensation and perception from the point of view of an objective psychology.

(1) Types of stimulus-response relations are considered as they relate to the following topics: (*a*) introspective or phenomenological descriptions, (*b*) psychophysical discriminations, (*c*) absolute judgments, and (*d*) thresholds and perceptual functions. Only (*b*) and its derivative (*d*) usually involve measured aspects of both stimulus and response. In the case of (*c*), the response is usually not quantitative. In the case of experiments involving introspection or phenomenological description, neither stimulus nor response is quantitative.

(2) The terms *sensation* and *perception* probably do not refer to two different operationally specifiable concepts.

(3) Problems of terminology are considered. Such terms as *hue, brightness,* and *saturation* are examined and analyzed. It would probably be possible in a behavior science to dispense with such terms, but they may have the advantage of reducing cumbersomeness. Their meanings within a context of stimulus-response relations are considered.

(4) Some recurring problems of perception are described, with particular reference to the kinds of variables involved. It is stated that the two major directions of perceptual investigation involve studies of the effects on discriminations of (*a*) stimulus conditions and (*b*) past history parameters and conditions of the subject.

(5) Analysis of the classes of stimulus-response functions concerned in perception indicates that the study of the latter topic becomes coextensive with the general study of behavior.

REFERENCES

1. BERGMANN, G., & SPENCE, K. W. The logic of psychological measurement. *Psychol. Rev.,* 1944, **51**, 1–24.
2. GIBSON, J. *Perception of the visual world.* Cambridge, Mass.: Riverside Press, 1950.
3. GRAHAM, C. H. Psychophysics and behavior. *J. gen. Psychol.,* 1934, **10**, 299–310.
4. GRAHAM, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.,* 1950, **57**, 108–120.
5. GRAHAM, C. H. Visual perception. In S. S. Stevens (Ed.), *A handbook of experimental psychology.* New York: Wiley, 1951.
6. GRAHAM, C. H. Behavior and the psychophysical methods: An analysis of some recent experiments. *Psychol. Rev.,* 1952, **59**, 62–70.
7. HECHT, S., SHLAER, S., & PIRENNE, M. H. Energy, quanta, and vision. *J. gen. Physiol.,* 1942, **25**, 819–840.
8. HELMHOLTZ, H. *Treatise on physiological optics. III. The perceptions of vision.* Ithaca, N. Y.: Opt. Soc. Amer., 1925.
9. HUNTER, W. S. The subject's report. *Psychol. Rev.,* 1925, **32**, 153–170.
10. KATZ, D. *The world of colour.* London: Kegan Paul, Trench, Trubner, 1935.
11. KOFFKA, K. *Principles of gestalt psychology.* New York: Liveright, 1940.
12. KÖHLER, W. *Dynamics in psychology.* New York: Liveright, 1940.
13. MACLEOD, R. B. An experimental investigation of brightness constancy. *Arch. Psychol.,* 1932, **135**, 1–102.
14. MUELLER, C. G. Frequency of seeing functions for intensity discrimination at various levels of adapting intensity. *J. gen. Physiol.,* 1951, **34**, 463–474.
15. SKINNER, B. F. The operational analysis of psychological terms. *Psychol. Rev.,* 1945, **52**, 270–277.
16. TITCHENER, E. B. *A text-book of psychology.* New York: Macmillan, 1919.

(Received May 15, 1957)

Psychological Review
Vol. 65, No. 2, 1958

TOWARD A QUANTITATIVE DESCRIPTION OF LEARNING SET DATA

FRANK RESTLE[1]

Michigan State University[2]

When the same monkey, ape, or child is given hundreds of object-discrimination problems to solve, he gets better and better at mastering them. Controls in the basic experiments eliminate two attractive explanations of this improvement. The cumulative transfer of training is *not* based on cues common to and peculiar to the correct objects, for problems are made up by random pairings from a pool of mutually discriminable objects. The cumulative improvement is not likely to be based solely on habits of conforming to the procedure of the experiment. Monkeys typically receive months of preparatory training in moving objects, in picking up rewards, and in general "good behavior" before the learning-set (LS) experiment itself begins.

It would appear, to the naive observer, that monkeys in an LS experiment "get the idea" of solving discrimination problems. An adult human subject, trained in this fashion, would soon realize that of the two objects presented in any problem, one is rewarded and the other not. He would learn that when new objects appear, the correct response is unpredictable until after one information trial, but that, in subsequent trials with the same objects, reward is consistently given and predictable. The data suggest that monkeys also respond in this fashion, though they acquire this general understanding of the situation only very slowly.

If we had available only the observation that monkeys improve in successive discrimination problems, it would be tempting to try an explanation in terms of the simple processes characteristic of naive rats. The compounding of further detailed information about the performance of sophisticated monkeys and apes makes such a reduction seem almost impossibly difficult. But an attempt to formulate the intuitively sensible hypothesis that monkeys gain an abstract understanding of the situation puts a great strain on current theoretical resources. Cognitive discussions of learning and performance are validly criticized on the ground that they do not yield precise predictions regarding behavior and thus are not conveniently verifiable. In the present paper, an attempt is made to formulate a descriptive theory of LS formation and operation which centers on an assumption of abstract cognition, and yet yields exact quantitative predictions which can be, and in some examples below will be, compared with the details of monkey performance.

The theoretical aim is to summarize a variety of data, using a few assumptions. To accomplish this kind of parsimony the writer has found it expedient to violate another informal canon of learning theory: that processes invoked should be simple and familiar. It is assumed that monkeys use an abstract understanding of the pattern of an LS experiment, transcending the "stimulus-response" rubric familiar in most theories of learning.

---

[1] The author is indebted to H. F. Harlow, J. M. Warren, and A. J. Riopelle for many valuable criticisms and insights and for making a variety of unpublished findings available.

[2] Part of the research on which this paper is based was done while the author was in residence at the Center for Advanced Study in the Behavioral Sciences.

Since the present formulation grew out of attempts to extend an earlier theory of discrimination and choice learning (**10, 11, 12**) to LS experiments, it is not surprising that the LS theory contains the earlier theory as a special case. However, it would be wrong to consider this LS theory a deduction from the simple discrimination learning theory. The logical implication is from the LS to the simpler theory, not *vice-versa*.

EMPIRICAL SCOPE OF THE THEORY

The present theory is intended to describe the change in height, rate, and shape of within-problems learning curves with successive problems. While the general characteristics of the data are represented in the theory, complications arising from our ignorance about how the stimulus objects appear to monkeys, and indeed about the actual stimulus properties of the objects, make it difficult to determine whether or not the theory is correct. In addition, as we shall see, there is evidence that the theory is incomplete, leaving out one of the factors which result in improved performance.

The formulation may also be related to reversal learning-set, and the solution of problems like oddities, matching from sample, delayed reaction, etc. These other problems can be thought of as involving (for the monkey) abstractions like, but different from, those involved in ordinary discrimination LS. The theory has not been elaborated enough to permit detailed analysis of these problems, and the present paper will merely sketch the possibilities.

Most important, for quantitative evaluation of the theory, is its use in analyzing the performance of sophisticated monkeys on various perceptual problems. After extended training the monkey reaches a relatively stable state, in which his performance on a variety of tasks can be studied intensively. It happens that the equations describing performance in this stable state are quite simple. Since the theory purports to relate learning to perceptual characteristics of the situation, it is possible to frame some interesting, if oversimplified, perceptual hypotheses, and then use the theory to evaluate them. Though these tests of the theory are separate from a large part of the general apparatus of the theory, and involve extraneous hypotheses about the stimuli, they are helpful in evaluating the theory because the interpretation of the data hinges on the crucial assumption of the whole structure—the relation between the perceptual situation and the rate of learning.

THEORY OF LEARNING SETS

The fundamental assumptions of the theory are: (*a*) Solving a discrimination problem involves two processes: conditioning valid cues to the correct response, and adapting or suppressing invalid cues. (*b*) These processes can be quantitatively described as the cumulative effects of a random process. (*c*) The rates of these processes (that is, the proportion completed on each trial) depend on the mean validity of cues present. (*d*) The probability of a correct response is the proportion of functional (unadapted) cues conditioned to the correct response. These assumptions have been stated in earlier papers (**10, 11, 12**).

In learning-set experiments involving higher primates, there exist certain special cues, which we shall term "type-*a*" cues, which are valid (consistently reinforced) and common to all problems of the experiment, remaining constant when the particular stimulus objects are changed. These cues are present on every trial except the first (information) trial of each problem, and they depend on previous reinforcements. In effect, a monkey responding to a type-*a* cue is responding to "the object which has been reinforced." He is responding to the property of having been reinforced, as distinguished from other properties such as size, color, spatial shape, arrangement of parts, etc.[3] These latter concrete properties give rise to type-*b* cues, which are valid (consistently reinforced) during a given problem, but which are not valid in the long run. There are other cues in the situation, arising from the apparatus, certain characteristics of the procedure, and, most important, from within the animal itself. These cues, termed type-*c*, are not reinforced consistently at any time, and thus are invalid.

Indications are that in the monkey, type-*a* cues make up a very small proportion of the cues the animal can use. Type-*b* make up a much larger proportion, about one-third of the cues, and type-*c* cues make up the rest, nearly two-thirds, in ordinary object-discrimination problems of the Wisconsin type (**4**). During any one problem, type-*a* and type-*b* cues become conditioned to the correct response and type-*c* cues become adapted, so that eventually the animal can attain perfect performance on the problem. But slowly, as more and more problems accumulate, type-*c* cues become adapted permanently, so that they have no effect on performance. Learning within a problem is then faster, since it need not wait upon the adaptation of the type-*c* cues. Then, even more gradually, type-*b* cues become adapted,[4] since they are not as valid as the type-*a* cues. When this process is nearly completed, the animal is responding in terms of the abstract type-*a* cues only. Since, by theory, these cues do not change from problem to problem, the animal does not learn new problems, but merely applies his previous knowledge again and again. As soon as one of the objects gains the property of "having been reinforced," through the one information trial, the animal responds perfectly and continues his accurate response for the rest of that problem.

Thus, this theory describes the fact that monkeys and higher primates (including children) become capable of one-trial solution of object-discrimination problems (**4, 6**). The accomplishment of this skill is described as a gradual process, again in agreement with the facts (**4, 6**). But more precise evaluation of the ideas depends on a more specific statement of the theory, in quantitative terms. We now turn to this problem.

The quantitative assumptions of this kind of discrimination learning theory are stated elsewhere (**10, 11, 12**). The present theory is identical with the most recent statement of these assumptions (**11**), except for a refinement of the concept of adaptation. Only a very brief resumé, as needed for this paper, is given here.

Any cue, *k*, is conditioned to one

---

[3] Verbal statements of the principle, by young children, are described by Hayes, Thompson, and Hayes (**6**).

[4] Riopelle has stated, and supported by experiment, the idea that highly sophisticated monkeys are not bound to concrete characteristics of the objects. In particular, he showed that such experienced monkeys rapidly forget previous experiences with particular objects (**13**). Limitations on the asymptotic performance of cats have been attributed by Warren to inability to go beyond the particular stimuli (personal communication, 1957).

or the other of the alternative responses. The probability that k will become conditioned to the reinforced response is θ, the mean validity of cues present. If K is the total set of cues present, V_k is the validity and θ_k the weight of cue k, then

$$\theta = \sum V_k\theta_k (k \text{ in } K) \qquad [1]$$

If $F_{(k,n)}$ is the probability that cue k is conditioned to the correct response on trial n, and the correct response is rewarded, then

$$F_{(k,n+1)} = F_{(k,n)}(1-\theta) + \theta \qquad [2]$$

whereas if the other (incorrect) response is reinforced,

$$F_{(k,n+1)} = F_{(k,n)}(1-\theta) \qquad [3]$$

is the correct response is reinforced for a block of j trials, $n+1$, $n+2$, $\ldots, n+j$,

$$F_{(k,n+j)} = 1 - [1 - F_{(k,n)}](1-\theta)^j \qquad [4]$$

This function, and its relation to a cumulative random process, are explained by Estes and Straughan (**2**).

In applications, we shall assume that the type-*a* cues are consistently reinforced on every trial except the first of each problem. If, at a certain trial of the experiment, there have been N such trials,

$$F_{(a,N)} = 1 - [1 - F_{(a,1)}](1-\theta)^{N-1} \qquad (a = \text{a type-}a \text{ cue}) \qquad [5]$$

Since these experiments involve thousands of trials, and θ is sizeable (at least over .10 in most applications), $F_{(a,N)}$ soon approaches 1.

We shall assume that type-*b* cues are reshuffled at each new problem, so that at the first trial of each problem some unknown proportion of them are conditioned to the correct response and the rest to the wrong response. Learning then progresses regularly during the problem. It is convenient to describe a trial as the nth trial of the current problem. Then

$$F_{(b,n)} = 1 - [1 - F_{(b,1)}](1-\theta)^{n-1} \qquad (b = \text{a type-}b \text{ cue}) \qquad [6]$$

Because the conditioning depends both on the total noninformation trials (N) and the number of trials on the current problem (n), we shall write $F_{(k,N,n)}$ as the probability that cue k is conditioned to the correct response at trial N, n, the Nth noninformational trial of the experiment, the nth in the current problem.

Since type-*c* cues are reinforced at random throughout the experiment, they cannot become conditioned consistently to the correct response. It is convenient for present purposes merely to state that

$$F_{(c,N,n)} = .5 \qquad (c = \text{a type-}c \text{ cue}) \qquad [7]$$

Adaptation of cues presents a somewhat more complicated problem, especially since some recent results of Hammer (**3**) have shed new light on the nature of learned perceptual selectivity. Previously, it has been stated that invalid cues become "adapted," suppressed, or nonfunctional (**10, 11, 12**). But Hammer showed that if certain invalid cues are adapted in the presence of some valid ones, and then the valid cues are changed, the invalid cues recover their function and are "de-adapted." Or, at least, there is no positive transfer of training. This suggests that perceptual selectivity is a more delicate process than mere adaptation of invalid cues. It appears that invalid cues are adapted *with respect to* valid cues present, and their status as nonfunctional depends on the presence of the valid cues. To say that invalid cue i is adapted with respect to valid cue v is to say that, when both are present, cue i has no effect on performance. If cue i is adapted with respect to several valid cues, then it is nonfunctional whenever it appears along with any of the valid cues.

The probability that cue i will be adapted with respect to cue v depends on the relative weight of cue v in the whole set, and also on the difference between i and v in validity. Specifically, we may write this probability as $\phi_{(i/v)}$. Letting θ_v be the relative weight of cue v, and V_v and V_i be the validities of cues v and i, then

$$\phi_{(i/v)} = \theta_v(V_v - V_i) \qquad [8]$$

Letting $A_{(i/v,N,n)}$ be the probability that cue i is adapted with respect to cue v at trial N, n, we have

$$A_{(i/v,N+1,n+1)} = A_{(i/v,N,n)} \times [1 - \phi_{(i/v)}] + \phi_{(i/v)} \qquad [9]$$

This formulation is a refinement of a previous statement regarding adaptation of cues under conditions of probable reinforcement (**11**).

The significance of the idea that cues are adapted with respect to specific other cues appears when we consider the adaptation of type-*c* cues in learning-set experiments. Within each problem, if it is solved, all type-*c* cues are adapted. But this seems to have little immediate effect on the next problem. The reason is that type-*c* cues are in the main adapted with respect to type-*b* cues, which are changed for the next problem. When this change is made, the type-*c* cues are released at once from adaptation, and have to be adapted all over again. When type-*c* cues happen to become adapted to type-*a* cues the adaptation is permanent, because the type-*a* cues do not change. Thus, within each problem the type-*c* cues are fairly rapidly adapted, but over many problems they gradually become permanently adapted.

Reducing these observations to difference equations, we have

$$A_{(a,N,n)} = 0 \qquad [10]$$

Type-*a* cues are not adapted because there are no other, more valid cues present with respect to which they can be adapted. Type-*b* cues have a certain validity, V_b, which is less than 1 because such concrete cues are not consistently reinforced over the entire experiment. Type-*b* cues can be adapted only with respect to type-*a* cues, so we may write $\phi_{(b/a)} = \theta_a(V_a - V_b)$. Assuming that $V_a = 1$, since type-*a* cues are consistently reinforced, we have

$$A_{(b/a,N+1,n+1)} = A_{(b/a,N,n)}[1 - \theta_a \times (1 - V_b)] + \theta_a(1 - V_b) \qquad [11]$$

Solving Equation 11 involves a conceptual problem. If a certain type-*b* cue is present only in some problems (as must be expected from the structure of a LS experiment), then the over-all probability that a certain cue k_b is adapted on trial n becomes $r\phi_{(b/a)}$, where r is the probability that k_b will be present on a given trial. For present purposes it is not feasible to estimate r, and we shall assume that $r = 1$. We are assuming that the several problems are made up by reshuffling a single set of type-*b* cues. This is certainly not exactly true. However, Riopelle (**15**) trained monkeys on LS with a minimal number of stimulus objects (that is, with r near 1), and then transferred the animals to ordinary LS (with r presumably lower). The absence of any consistent decrement in transfer indicates that LS is quite general, and that r is approximately 1. In "concept-formation" experiments with humans (**7**, pp. 232–241), perfect performance is often maintained despite continuous changes in the particular stimuli presented. This again indicates that variations in stimulus objects may be more like reshuffling than one would think.

If essentially the same type-b cues appear in most problems, we should apply Equation 11 on every trial where type-a cues appear. The solution is

$$A_{(b/a,N,n)} = 1 - [1 - \theta_a(1 - V_b)]^{N-1} \quad [12]$$

Adaptation of type-c cues is somewhat more complex, since they can be adapted with respect to either type-a or type-b cues. The probability that any type-c cue will be adapted to a type-a cue on trial n, given that it is not already adapted to a type-a cue, is θ_a, since $V_a = 1$ and $V_c = 0$. The probability that the type-c cue will be adapted with respect to a type-b cue, given that it is not already adapted to a type-b cue, is $V_b\theta_b$. This statement may be related to a "cue-sampling" notion by imagining that a given cue can be adapted with respect to both type-a and type-b cues; however, on a given trial, only one type of adaptation can occur. A cue cannot go from the state of being adapted with respect to neither a nor b cues into the state of being adapted with respect to both, in the space of a single trial.

With this formulation we get the difference equations

$$A_{(c/a,N+1,n+1)} = A_{(c/a,N,n)} \times (1 - \theta_a) + \theta_a$$

and [13]

$$A_{(c/b,N+1,n+1)} = A_{(c/b,N,n)} \times (1 - V_b\theta_b) + V_b\theta_b$$

Then the probability that c is adapted with respect to at least one of the two types of cues, $A_{(c/a+b,N,n)}$, follows the difference equation

$$A_{(c/a+b,N+1,n+1)} = A_{(c/a+b,N,n)} \times (1 - \theta_a - V_b\theta_b) + (\theta_a + V_b\theta_b)$$

so that

$$A_{(c/a+b,N+j,j)} = 1 - [1 - A_{(c/a+b,N+1,1)}] \times (1 - \theta_a - V_b\theta_b)^{j-1} \quad [14]$$

Earlier discussion of adaptation suggests that, when a new problem is introduced, type-c cues adapted to type-b cues only become released from adaptation, whereas cues adapted to type-a cues remain adapted. Then

$$A_{(c/b,N+1,1)} = 0$$

so that

$$A_{(c/a+b,N+1,1)} = A_{(c/a,N+1,1)}$$

which from Equation 13 can be shown to be

$$A_{(c/a,N+1,1)} = 1 - [1 - A_{(c/a,1,1)}](1 - \theta_a)^N$$

No cues are adapted at the beginning of the whole LS experiment. Thus,

$$A_{(c/a,1,1)} = 0$$
$$A_{(c/a,N+1,n)} = 1 - (1 - \theta_a)^N$$

Substituting this expression into Equation 14, we have, finally,

$$A_{(c/a+b,N+j,j)} = 1 - (1 - \theta_a)^{N-1} \times (1 - \theta_a - V_b\theta_b)^{j-1} \quad [15]$$

According to the discrimination theory (**11**), the probability of a correct response is the proportion of unadapted cues conditioned to the correct response. This probability, which we here call $p_{(N,n)}$, is given by

$$p_{(N,n)} = \frac{\sum F_{(k,N,n)}[1 - A_{(k,N,n)}]}{\sum[1 - A_{(k,N,n)}]} \quad \text{(all cues } k\text{)} \quad [16]$$

which may be expanded using equations 5, 6, 7, 10, 11, and 15, as

$$p_{(N,n)} = \frac{\theta_a[1 - (1 - F_{(a,1,1)})(1 - \theta)^{N-1}] + \theta_b[1 - (1 - F_{(b,N,1)})(1 - \theta)^{n-1}] \times [(1 - \theta_a(1 - V_b)]^{N-1} + .5\,\theta_c(1 - \theta_a)^{N-1}(1 - \theta_b V_b)^{n-1}}{\theta_a + \theta_b[(1 - \theta_a(1 - V_b)]^{N-1} + \theta_c(1 - \theta_a)^{N-1}(1 - \theta_b V_b)^{n-1}} \qquad [17]$$

While Equation 17 is large and complicated in form, it contains only three main independent parameters, each of which may be manipulated separately. The parameters are θ_a, the proportion of type-*a* cues; θ_b, the proportion of type-*b* cues; and V_b, the validity of the type-*b* cues. The other parameters, θ_c and θ, are simply defined in terms of the three independent ones, for $\theta_c = 1 - \theta_a - \theta_b$, and $\theta = \theta_a + \theta_b V_b$. If data on many problems are to be averaged, we may let $F_{(a,1,1)}$ and $F_{(b,N,1)}$ be one-half.

Equation 17 purports to give the probability of a correct response on any trial of a learning-set experiment. Tests of the theory will use Equation 17 and certain simplified special cases of it.

APPLICATIONS OF THE THEORY TO DATA

The first question is whether Equation 17 properly describes the set of learning curves arising during the formation of learning sets. The complicated form of the equation precludes the use of standard curve-fitting methods to test the theory. The more refined stochastic-model methods are equally inaccessible, for the same reason. The only method of evaluating the theory is to evaluate the parameters θ_a, θ_b, and V_b by a method of approximations, and set Equation 17, with those approximate parameters, alongside empirical performance curves. Harlow (4) has published learning-set functions for a group of monkeys learning object discriminations. To a fair approximation, these data are fitted by Equation 17 with

$$\theta_a = .0075$$
$$\theta_b = .350$$
$$V_b = .70$$

While these parameters are clearly only approximate, the correspondence between Equation 17 and the data is quite good, as shown in Fig. 1.

Inspection of Fig. 1 reveals that the changes in height, slope, and shape of the within-problems curves are followed by the theoretical function.

However, a second computation was performed on pattern-discrimination data published by Harlow and Warren (**5**). Here, it was impossible to get a good fit, because the rate of within-problems learning, based mainly on θ_b, seems to increase during learning for the monkeys. The theory tacitly assumes that θ_b, the proportion of type-*b* cues, is constant throughout the experiment. It appears that the monkey's attention is initially directed to the edges of the *objects* presented, and thus away from the *patterns* drawn on those objects, since the patterns are centered on the objects. The monkeys probably acquire some sort of "observing response" (**23**), directing their visual fixations toward the centers of the objects. If this should happen, it would increase the relative weight of the pattern cues (type-*b* cues) so that θ_b would increase during the experiment.

The success of Equation 17 for object-discrimination, where no special

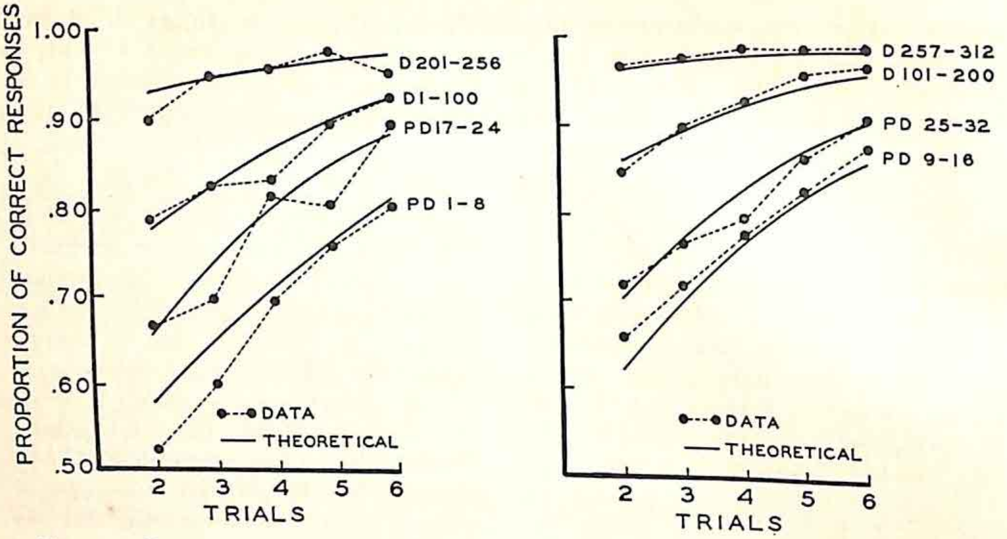


Fig. 1. Comparison of theoretical learning set function, Eq. 17, with $\theta_a = .0075$, $\theta_b = .350$, $V_b = .70$, and some data from Harlow (4). Alternate sequences of problems are shown in two panels to prevent crowding.

visual habits are required, and its failure for pattern discrimination, where the monkey must overcome a pre-experimental predilection, indicate that the present theory might be used to isolate the formation and effects of observing responses.[5]

In ordinary learning set, the type-*a* cues are conceived to be abstract conditional discrimination cues—"abstract" because the animal can learn to respond independently of the concrete, type-*b* cues present and "conditional" because the object chosen is contingent on previous reinforcements or nonreinforcements of those objects. In discrimination learning-set experiments, response to such abstract, conditional cues is consistently reinforced. In discrimination-*reversal* experiments, the object recently reinforced becomes the wrong object, abruptly, in the middle of the experiment. The monkey must learn a new contingency, for he must learn to reverse objects to the cue of an unexpected nonreinforcement. That reversals can be cued, by controlled stimuli, has been demonstrated in an experiment by Riopelle and Copelan (**16**). The mechanism to be suggested here for reversal learning set is essentially the one they were analyzing —another type of cue, which might be called type-a_r, which in reversal learning can come to supercede regular type-*a* cues. Since, right after reversal, the animal would have no clear cue based only on type-*a* cues (which object, after all, is the one which "has been reinforced"?—both objects have been reinforced), it is reasonable to believe that some more fitting and precise abstract cue may arise to control behavior.

The present theory relates rather directly to Meyer's equations for discrimination and reversal learning (**8**, **9**). Meyer noted that within-problems learning curves from fairly sophisticated animals are exponential in form, with a constant rate but a gradually increasing origin. If, in Equation 17 above, we assume that the type-*c* cues are almost all adapted (i.e., $A_{(c/a+b,N,n)}$ is almost 1) and the type-*a* cues are all conditioned to the correct response ($F_{(a,N,n)} = 1$), then

---

[5] A theory of learning set, which depends mainly on observing responses as the mechanism, is possible. In the present theory, observing responses are considered possible special influences on behavior, but do not play a central explanatory role.

$$p_{(N,n)} = \frac{\theta_a + \theta_b[1 - (1 - F_{(b,N,1)})(1 - \theta)^{n-1}][1 - \theta_a(1 - V_b)]^{N-1}}{\theta_a + \theta_b[1 - \theta_a(1 - V_b)]^{N-1}} \quad [18]$$

Since the adaptation of type-*b* cues is very slow relative to the conditioning of them, a given within-problems learning curve would be very closely approximated by considering $[1 - \theta_a(1 - V_b)]^{N-1}$ a constant. Elementary algebra shows that in this case

$$p_{(N,n)} = 1 - \beta(1 - \theta)^{n-1} \quad [19]$$

where β depends on the degree of adaptation of type-*b* cues and also on the animal's original preference for one object or the other. If almost all type-*b* cues are adapted, β becomes quite small. Since Equations 17, 18, and 19 all include type-*a* cues as present, of course, we must assume that we are talking, not about an information trial, but some later trial (Trial 2, 3, etc.) in the problem. Equation 19 shows that the within-problems learning curve, for a fairly sophisticated animal, will be nearly exponential in shape. It is also worth noting that the rate of this exponential curve depends on θ, which is a constant except for possible observing responses. Meyer's observation was that within-problems learning curves were roughly exponential, with a constant rate. His rate, estimated from his own data, was .251, which corresponds closely to the rate we selected for Harlow's data. In Harlow's data $\theta = \theta_a + \theta_b V_b$ was $.0070 + .350(.70) = .2520$. The gradual increase in the intercept, or performance on the first postinformation trial, according to the present theory depends on adaptation of type-*b* cues, and a corresponding decrease in β.

In an oddities problem, the monkey must select whichever object, of three or more, is the odd one. His performance will be no better than chance if he selects the object which last was reinforced, since that may not be odd. Monkeys can solve such problems, presumably on the basis of an appropriate abstract cue, but the formation of a discrimination learning set seems to be of little help, and perhaps a hindrance. This observation checks with the theoretical interpretation.

It should be clear that further analysis of data from experiments on reversal learning, oddities, and so forth, is needed. The general method of formulation suggested by this theory is to hypothesize a kind of type-*a* cue corresponding to the procedure employed, taking account of possible cues the monkey could use to solve the problem. These conditional, abstract cues can be evaluated as to their relative weight (θ_a), using Equation 17 or reasonable modifications thereof. Interrelations between the various abstract cues are sometimes apparent from inspection. A number of tests of the theory are possible along these lines, and may be worth pursuit.

Equation 19 is the basis of other tests of the theory. Using this equation, we can analyze the performance of monkeys which are fairly sophisticated. The important idea is that in Equation 19, the parameter θ depends mainly on the proportion of type-*b* cues. This is so because $\theta = \theta_a + \theta_b V_b$, and θ_a in the monkey seems to be very small compared with θ_b. To an approximation, then, $\theta \cong \theta_b V_b$. It should be recalled that type-*b* cues are concrete, differential cues. Differences between the discriminanda in color, size, form, or other properties are represented as sets of cues of type *b*. The relative amount of valid differential cues theoretically controls θ in Equation 19.

An example of the use of Equation 19 is in an experiment on additivity of cues, performed by Warren (**21**). The same monkeys solved a variety of pattern discriminations in which the differential cues were form (F), color (C), size (S), or combinations of two characteristics (F + C, F + S, C + S), or combinations of all three dimensions (objects differing in F + C + S). At the beginning of testing the animals were fairly sophisticated.

Assuming that β would be the same for all the various problems, it was possible to estimate the parameters for the F, C, and S problems. The parameters selected were

$$\beta = .35$$
$$\theta_F V_F = .089$$
$$\theta_C V_C = .221$$
$$\theta_S V_S = .067$$

These parameters were selected by visual curve-fitting, β depending on the intercept and θV depending on the learning rate.

Assuming that in these problems, as in object discrimination, the validity of type-*b* cues is about .70, so that $V_F = V_C = V_S = .70$, the above equations can be solved for θ_F, θ_C, and θ_S, the proportions of type-*b* cues in the three types of problems. The values obtained are

$$\theta_F = .127$$
$$\theta_C = .316$$
$$\theta_S = .096$$

From these parameters, we seek to predict performance on multiple-cue problems, that is, the F + C, F + S, C + S, and F + C + S problems. We first note that if we let I be the number (or weight) of type-*c* cues, F be the weight of form cues, C be the weight of color cues, and S be the weight of size cues, we have the relationships

$$\theta_F = F/F + I$$

or

$$F = I\theta_F/(1 - \theta_F) = .146I$$
$$C = I\theta_C/(1 - \theta_C) = .462I$$
$$S = I\theta_S/(1 - \theta_S) = .106I$$

We now note that if form, color, and size cues are additive,

$$\theta_{F+C} = F + C/F + C + I = .146I + .462I/.146I + .462I + I = .379$$

Similarly,

$$\theta_{F+S} = .146I + .106I/.146I + .106I + I = .201$$
$$\theta_{C+S} = .362$$
$$\theta_{F+C+S} = .416$$

By assumption, V in all these cases is .70, so we may at once compute the predicted learning rates for these combined-cues problems. They are

$$\theta_{F+C} V_{F+C} = .265$$
$$\theta_{F+S} V_{F+S} = .141$$
$$\theta_{C+S} V_{C+S} = .253$$
$$\theta_{F+C+S} V_{F+C+S} = .292$$

In all cases, we shall continue to assume that $\beta = .35$. Now, we have computed complete predictions for the four multiple-cue problems, without reference to the data of such groups. For example, the learning curve for $F + C$ problems should be

$$p_{(N,n)} = 1 - .35(1 - .265)^{n-1}$$

for all trials, n, except the first. This prediction can be compared with the

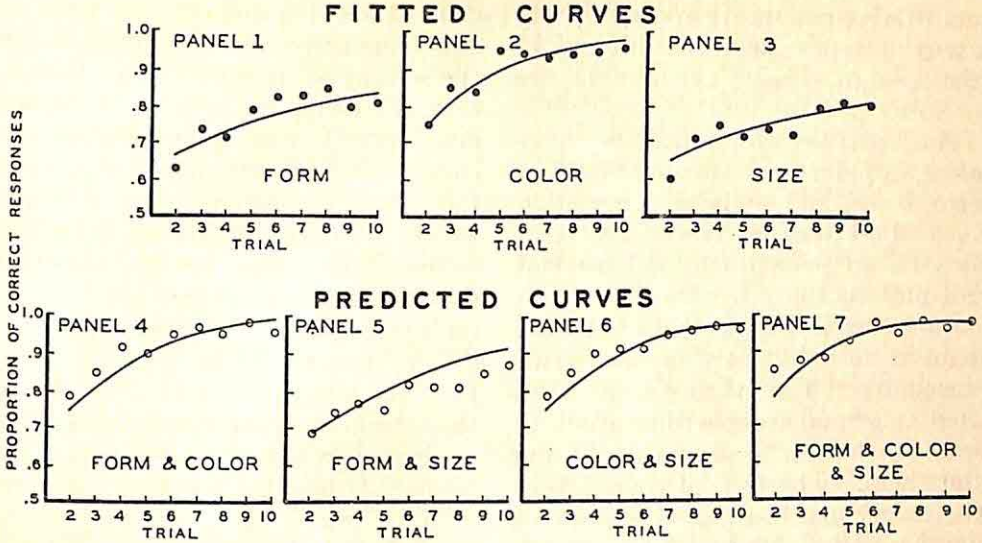


FIG. 2. Fitted within-problems learning curves for single cues, and predicted curves for combined cues, in Warren's experiment on cue-additivity. Panels 1, 2, and 3 show curves fitted by approximation. Panels 4, 5, 6, and 7 show a priori curves. All curves are from Equation 19. For parameters and rationale, see text. Data from Warren (21).

performance of the animals on the F + C problems.

The fitted curves for the three separate cue problems, and the predicted curves for the combined-cues problems, are plotted in Fig. 2. Inspection shows that the predictions are quite successful, almost as good as the fitted curves with which we started.

The fact that the empirical curves do not seem to be the same shape as the theoretical curves may be due, at least in part, to the fact that empirical curves represent averages over a variety of problems of the same type which differ in difficulty, and over the performance of different monkeys which differed significantly. The average of a set of exponential functions, with different rates, is not itself an exponential function (1). The estimates of parameters are at best rough approximations, but it is encouraging to see that the predictions are at least approximately correct, even despite the irregularities of detail.

Warren also has studied performance of monkeys first trained on an F + C + S problem, then retrained with some of the same cues present but some removed (**22**). A drop in performance was observed whenever any cue was removed. The relative amount of drop caused by removal of various cues, or combinations of cues, checks well with our observed parameters. For example, color had the greatest effect, more than form and size together, and form was slightly more important than size. The stimuli used in this experiment were quite similar to, though not identical with, those used in the experiment on additivity (**21**).

Theoretically, removal of part of the type-*b* cues would result in recovery from adaptation of certain type-*c* cues which had been adapted only with respect to the cues removed. The greater the proportion of cues removed, the more invalid cues would recover, and the more performance would drop back toward chance. The quantitative results are approximately in accord with expectation, though in this case no specific calculations are possible.

The above computations were based, in part, on the assumption that form, color, and shape cues are additive. This means, essentially, that the cues are perceptually independent and have nothing in common. This assumption is supplementary to the main theory, but in this case seems reasonable. The theory has been used as a framework within which to state such special hypotheses, and test them. The test, of course, is of the theory *and* the special hypothesis jointly, and failures of such predictions may be attributed either to failure of the theory or of the special hypothesis. This weakens the power of such experiments to verify the theory. Nevertheless, one must keep in mind the distinction between the basic theory and the special assumptions, lest the theory be prematurely discarded because it is combined with an ill-conceived special assumption. It is by no means sure that color, form, and size are additive cues. Brightness and size are probably not independent cues, since they have the common factor of total flux of light.

A more tenuous perceptual hypothesis has also been tested, using the present theory. The results are interesting, if inconclusive. Warren (**20**) studied the rate of solution of form discriminations as a function of the size of forms used. In each problem the two forms were of equal size, but some problems used small forms and other problems used larger forms. On a 3″ × 3″ object, forms were of 0.9, 1.8, 2.7, 3.6, 4.5, and 5.4 sq. in. areas. The larger forms were more rapidly mastered, indicating that larger forms give rise to more valid cues.

The following quantitative hypothesis was considered. Suppose that the weight or number of cues arising from the forms is proportional to the area of the forms, while the weight of cues arising from sources other than the forms is constant. Let I be the weight of cues not arising from the forms. The I cues are type-c, since only the forms are differential in Warren's experiment. The weight of cues arising from the forms is proportional to s, the size of the forms. Some of these form-cues are valid type-b cues, arising from differences between the forms. Other form-based cues are invalid type-c cues, arising from nondifferential characteristics of the forms. Let sB be the weight of the type-b cues arising from forms of size s, and sC be the weight of type-c cues arising from forms of size s. The proportion of type-b cues, when forms are of size s, will be

$$\theta_s = \frac{sB}{sB + sC + I} \qquad [20]$$

according to the hypothesis.

We may estimate θ_s for each size of forms employed, and see whether Equation 20 fits the obtained estimates. Since only the *relative* weights

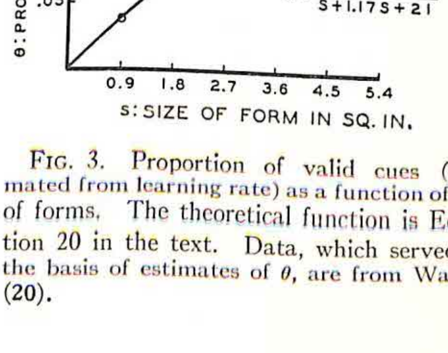


FIG. 3. Proportion of valid cues (estimated from learning rate) as a function of size of forms. The theoretical function is Equation 20 in the text. Data, which served as the basis of estimates of θ, are from Warren (**20**).

of type-b and type-c cues are important, we may take B as our unit. Equation 20 was fitted to estimates from Warren's data, with $B = 1$, $C = 1.17$, and $I = 21.0$, and the correspondence was found to be close (see Fig. 3).

This test of the theory is considered inconclusive, because the perceptual hypothesis is doubtful. It has been found that the saliency of a form on a background for the monkey depends on whether the form is near the edges of the object or not (**19**; and personal communication from A. J. Riopelle). Since increasing the size of a form which is centered on an object of fixed size brings the form nearer the edge of the object, it seems likely that a more complicated perceptual hypothesis will be needed.

DISCUSSION

This paper has expounded a theory of the formation and operation of "learning sets" in monkeys. Using an earlier theory of discriminative learning as a basis, it has here been assumed that monkeys are capable of responding to abstract, contingent cues. Monkeys, with extended experience, can transcend the concrete stimulus characteristics which ordinarily form the basis of discriminative responses, and respond to conditional cues such as "the object which was correct on the previous trial," etc. These contingent cues are analogous to the cues, whatever they may be, in conditional discriminations—one part of the compound being the characteristics of the object, the other part being previous reinforcements given and perceived by the animal.

This theory emphasized the *cue* value of rewards. Some motivational component of reinforcement is essential, even in the present theory, so that we can define "validity" and "the correct response," but rewards actually must have two effects. One, the traditional reinforcement effect, is as an agent to strengthen certain responses. The other is a cue-value, since food rewards are themselves perceivable cues.

The importance of rewards as cues has been demonstrated in a series of papers by Riopelle and his associates (**14, 17, 18**), who have shown that highly efficient performance can be maintained even when no reward is given on the first trial—so long as some informational cue, such as a marble, is used to indicate which object will be reinforced. In the present terminology, they have shown that type-a cues, which are usually contingent on previous rewards, can as well be contingent on other things, such as marbles, which have the same place in the procedure as rewards. It is, of course, doubtful that monkeys would long perform effectively if no rewards at all are given for correct responses, since the reinforcing (strengthening) property of rewards is essential. But on Trial 1, in a problem with a sophisticated animal, the reward seems to function mainly as information, and one signal is about as good as another. In fact, since raisins seem to capture the animal's whole attention and detract from observation of the object, it was found that the marble signal was sometimes more efficient than the reward.

It should be evident that the basic concepts of the present formulation are not original with the writer, but instead are drawn from the discussions of writers like Harlow, Meyer, Riopelle, and Warren. The purpose of this study has been to develop a quantitative framework for ideas about learning sets which already existed in the literature. In the present formulation these ideas have of course been modified, and the formal relation-

Acc. No. ....
Dated ....
DAVID HARE TRAINING COLLEGE

ships, such as the mechanisms of conditioning and adaptation, and the relationship between cue-validity and rates of learning, are essentially those which were stated in previous papers by the present writer. Apparently, despite the changes and additions, there is some correspondence between the present theory and the theoretical interpretations of learning-set experimenters.[6]

The present theory is sharply limited as to the variables considered. The main factors in performance discussed are θ_a, the relative weight of the abstract cues, θ_b, the relative weight of the concrete differential cues, and V_b, the validity of the type-*b* cues. Only very simple perceptual hypotheses were used in testing the theory. The main important factor which appears to be missing is the possible appearance of observing responses. Other important factors are presumably missing, but have not yet been identified.

Tests of the theory are at present limited to a few fairly obvious and simple computations. Detailed analysis of the large body of experimental data, as indicated by the present theory, may be useful in systematizing and clarifying some of the findings already available. At the same time, extension of the theory to formal statements about reversal, oddities, and other such problems is indicated.

It should be noted that the present theory does not "explain" learning sets, in the sense of reducing the observed behavior to some simpler type of performance. Learning-set performance has not, for example, been shown to be merely another example of S-R association, or innate organization. The present theory is descriptive, attempting to write equations which enable one to predict performance in certain experimental settings. The equations do not arise from simple first principles, but are in a real sense *ad hoc*, with the restriction that they reduce to an earlier theory of discrimination learning for a subject's first problem, or for a subject bereft of type-*a* cues.

Two proper criticisms of the present theory may be mentioned here. One is that the whole formulation hinges on the existence of type-*a* cues, but there is no direct evidence that such cues are present. To this the only answer is that the cues are hypothetical, and any hypothesis is acceptable until it is either disproved or replaced by a better one. A deeper criticism would point out that type-*a* cues are not well described. Several crucial questions have been left unanswered: Are such cues innate or acquired, and if acquired, when are they acquired? How, exactly, are such cues to be described, since they seem to defy the usual sort of S-R analysis? Does the existence of type-*a* cues depend on the phylogenetic level, neural or receptor endowment, or "intelligence" of the subject employed? What kind of neural basis would be required? The only consequence of the presence of type-*a* cues, as the theory stands, is their effect on the performance of sophisticated monkeys—but certainly such performance might arise in any of a variety of ways. Answers to such questions can be obtained only by further investigations. It is hoped that the theory in its present form can take a place as one of the tools in the analysis of the mind of the monkey. Whether or not the tool is useful can be decided only by test.

---

[6] At least, discussions with Harlow, Riopelle, and Warren have indicated that such agreements exist.

REFERENCES

1. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
2. ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
3. HAMMER, M. The role of irrelevant stimuli in human discrimination learning. *J. exp. Psychol.*, 1955, **50**, 47–50.
4. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
5. HARLOW, H. F., & WARREN, J. M. Formation and transfer of discrimination learning sets. *J. comp. physiol. Psychol.*, 1952, **45**, 482–489.
6. HAYES, K. J., THOMPSON, R., & HAYES, C. Discrimination learning set in chimpanzees. *J. comp. physiol. Psychol.*, 1953, **46**, 99–104.
7. JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.
8. MEYER, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.
9. MEYER, D. R. Intraproblem-interproblem relationships in learning by monkeys. *J. comp. physiol. Psychol.*, 1951, **44**, 162–167.
10. RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.
11. RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.
12. RESTLE, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, **64**, 217–228.
13. RIOPELLE, A. J. Transfer suppression and learning sets. *J. comp. physiol. Psychol.*, 1953, **46**, 108–114.
14. RIOPELLE, A. J. Rewards, preferences, and learning sets. *Psychol. Reports*, 1955, **1**, 167–173.
15. RIOPELLE, A. J. Learning sets from minimal stimuli. *J. exp. Psychol.*, 1955, **49**, 16–22.
16. RIOPELLE, A. J., & COPELAN, E. L. Discrimination reversal to a sign. *J. exp. Psychol.*, 1954, **48**, 143–145.
17. RIOPELLE, A. J., & FRANCISCO, E. W. Discrimination learning performance under different first-trial procedures. *J. comp. physiol. Psychol.*, 1955, **48**, 90–93.
18. RIOPELLE, A. J., FRANCISCO, E. W., & ADES, H. W. Differential first-trial procedures and discrimination learning performance. *J. comp. physiol. Psychol.*, 1954, **47**, 293–297.
19. WARREN, J. M. The influence of area and arrangement on visual pattern discimination by monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 231–236.
20. WARREN, J. M. Effect of geometrical regularity on visual form discrimination by monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 237–240.
21. WARREN, J. M. Additivity of cues in visual pattern discrimination by monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 484–486.
22. WARREN, J. M. Perceptual dominance in discrimination learning by monkeys. *J. comp. physiol. Psychol.*, 1954, **47**, 290–292.
23. WYCKOFF, L. B. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received May 21, 1957)

Psychological Review
Vol. 65, No. 2, 1958

TRIAL DESIGN IN HUMAN EXPERIMENTS[1]

CHARLES W. SLACK

Harvard University

With notable exceptions such as free operant behavior (12) and some pure phenomenological experiments (7), psychological experiments consist of trials. Trials constitute the "building blocks" out of which experiments are made, and that branch of the science which deals with the arrangement of numbers of trials under differing conditions and over *S*s is the study of experimental design. It tries to tell us how we can distribute trials or blocks of trials so that we may get the most out of an experiment.

In this paper the author will try to show that attention can profitably be paid to the *trial itself*, to its composition independent of the particular parameters and variables used. Although many or most of our experiments are made up of trials, the trials per se have not been studied for what they are, and the principles of experimental design have not been extended to cover what goes on "inside" the trial. Trials are the "molecules" out of which experiments are constructed, and the trials themselves can be considered as made up of elements which are the critical events—stimuli, responses, and instructions.

The aim of this paper is threefold: (*a*) To bring attention to the study of the design of trials as an important and integral branch of experimental design, to propose that the trial can profitably be studied in its own right as an adjunct to psychological theory and method; (*b*) to propose one possible model, an "event arrangement" model, as being fairly adequate to cover a number of interesting traditional and novel experiments with human *S*s; (*c*) to show, by example with this model, how models of trial design can be used to generate new experiments, to provide a framework within which we can discuss experimental results, and, specifically, to provide an operational way of describing correspondences between theoretically derived trial types and experimentally obtained similarities and differences in response measures.

If we are correct, trial design could emerge as a methodology particularly concerned with and suited to the needs of psychology and developed for those needs rather than borrowed from other sciences like agriculture. No other science has experiments made up of trials in the same sense as psychology has. In most psychological experiments a trial is not an arbitrary unit. Psychological experiments are sequences of events, and the fundamental nature of the process studied is more contingent upon some aspects of the sequence than others. We may move some things around, and not others. For example, in a simple size-constancy experiment (*S*s judge the height of different sized objects at a distance and do this over a number of trials), we may trade or eliminate trials or blocks of trials, we may change variables and parameters including instructions, we may run each trial on a different *S*, and still be studying something which we call perception or a "perceptual type" response. However, there is a limit to the smallness of the unit which we can tamper with and still be studying anything like perception. If we stimulate one *S* and get the response from another independent *S*, or if we put the response *before* the stimulus in each trial, we have changed the whole "quality" or "character" of the experiment, and other changes (notably in instructions) are necessitated.

---

[1] This investigation has been aided in part by a grant from the Foundations' Fund for Research in Psychiatry.

DEFINITIONS

There are two general groups of characteristics which make one psychological experiment different from another when they are both made up of trials. In the first group we have those *variables* which change from trial to trial and those *parameters* which change from experiment to experiment (experiment = block of trials); and, second, we have the ordering over time of the occurrence of events within the trial, between trials, and between blocks of trials or experiments. Thus, for example, time in the sense of interval belongs in the group of parameters and variables, but the ordering of events over time is a problem of experimental design and is nonparametric by nature. Traditional experimental design is concerned with arrangement effects *between* trials and *between* blocks of trials. The present model of trial design concerns itself with arrangement effects *within* trials.

The usefulness of trial design, like any other methodology, rests upon the definitions of its fundamental concepts. I have previously rendered an analytic proof of the impossibility of a watertight a priori distinction between stimulus and response (11). Complete definitions of the event-elements which make up a trial are impossible without a number of assumptions regarding the relationship between organism and environment. Thus the definition of the components of a trial can never be a purely operational procedure, if by "operational" we mean that the definition can be made before empirical data are obtained.

The distinction which we, as *E*s, make between stimulus and instructions, for example, is never really clear in any particular case (11) and, furthermore, in the case of multiple-event trials it is very difficult at times to state beforehand how many events of a particular sort we are really dealing with. As we shall see, however, the purposes of trial design do not really demand that we be able to give precise definitions of the components—only that we be able to distinguish them well enough to order them in the trial.

Trials. Trials are unit parts of experiments which are blocks of trials. A trial is the smallest consecutive sequence of event-elements which is repeatable and which contains at least one of each of the critical event-elements (stimulus, instruction, and response). A trial cannot be broken down any further or otherwise altered without altering the character of the experiment, i.e., without changing a perceptual type of experiment into a learning experiment.

The event-elements within a trial *refer to* each other more than they do to other elements outside the trial. Two event-elements are said to refer to each other (and hence belong in the same trial) provided a main effect of the experiment would be discovered by observing the relationship between those events. Thus, in an ISR trial the stimulus and the response within the trial are said to refer to each other, since the relationship between these two events is a main effect of a perception experiment, whereas the relationship between, say, the stimulus of this trial and the response of the preceding trial is not a perceptual effect.

Reference is not the same thing as cause and effect, however, since a response may intimately refer to a stimulus yet to come and may, therefore, be neither the cause nor the effect of that stimulus. In this case the response is a predictive response and might, according to some theory, be "caused" by a previous stimulus not a part of the present trial.

Events Which Compose Trials. Trial design concerns itself with the composition of the trial. In the present model composition means the combination and permutation of event-elements. The critical events which make up a trial are, according to the model, stimuli (S's), responses (R's), and instructions (I's). Although it may be an advantage in psychological theory to consider instructions as no different from stimuli, in designing trials there is great advantage in considering them as having properties somewhat different from stimuli.

S, R, and I are defined as critical changes in what we know to be ongoing processes. The notion of "change" or differential is necessary since the model is an event-arrangement model; and if an event is continuous throughout a trial or an experiment, it cannot be arranged or moved around. Otherwise, the use of S, R, and I does not differ in any important way from many of the most common usages in the literature.

Stimulus. S is a change within a trial in the stimulating conditions. It is important that reinforcements in learning trials are always S's, as are perceptual and auditory displays, test items, etc. S may be defined in either psychological or physical terms depending upon the demands of the situation or the bias of the investigator.

Response. We know that the organism is always responding in some sense or other. Certain changes or rates of change in the R process are interesting and relevant, whereas others are not. An R is a critical differential in the output of the organism during a trial.

Instruction. I's are a kind of "conditioning" or super-determining stimuli which have the psychological property of orienting or setting the organism for R's (or sometimes for S's) to come. I's refer to and must precede R's, since they must be not incapable of altering the R to which they refer. However, R's may precede S's, in which case they are predictive R's. I's are often assumed to be unequivocal, often their physical properties are more trivial than S's, and often they are given in written and spoken words. None of these last characteristics is necessary, however. Reinforcements are never considered as I's—even when verbally delivered. Hence the phrase "You have done very well," is, according to the model, an S.

THE EVENT-ARRANGEMENT MODEL

The event-arrangement model works in the following way: by combining and permuting the elements S, R, and I, and adhering to the restrictions, trials of varying complexity can be generated. The model provides a way of breaking down complex trials into simpler trials and of building up ideal trials out of component trials. All of this is done without directly considering the variables of the experiments. Consideration for variables like time can be introduced later. By deducing *all* of the possible trial varieties central to a particular problem, issues take on a new light. Trials are thought of as controls for one another, and answers to questions are possible.

If it is difficult for the reader to consider trials in the abstract without knowing what the variables are, he can be aided by the following device. Imagine a very simple experiment in which the S varies in two ways with two values to each way (long, short; red, black); the R can be one and only one of these four. The I tells the *S* to attend to one of the two aspects (length or color). Now if we run the trials in the order ISR, we are studying "perception," and if **S** varies randomly and **I** is constant ("What length is it?"), we can be sure that any correlation between **S** and **R** over repetitions of the trial is due to "perceptual" accuracy, and we can define perceptual accuracy by that correlation. It is a measure of "how much the *S* sees." If we run the trials **IRS**, we are studying learning, and if **I** is constant ("What length will it be?") and **S** varies in a learnable manner, we can be sure that any within-trial **S-R** correlation reflects learning—"how much the *S* has learned." If we run the trials **SIR**, we are studying something conceivably having to do with perception or memory more than learning, but how it differs from **ISR** is not obvious. However, **SIR** suggests that we should vary **I**, and if **S** is constant and **I** varies randomly (sometimes length, sometimes color), we might say that **I-R** correlation measures the degree to which the *S* can change set. With some adaptations which are obvious in each case, one can extend this example with its simple variables to the complex trials.

SINGLE EVENT TRIALS

There are six permutations of the three "gross" events **I**, **S**, and **R**. Three of these are not consistent with the restriction that **I**'s must *refer to* an **R** within the same trial and must "set" the organism for that **R**.[2] To do this, **I** must precede **R**.

By elimination we have three consistent single event trials: **ISR**, **IRS**, and **SIR**. These three paradigms are all used fairly frequently in psychological experiments; as a matter of fact, together with the **ISRS** trial discussed later, they practically exhaust the presently explored trial varieties.

The **ISR** paradigm is the typical perception, discrimination, and sensation paradigm. It comes to us from psychophysics and is incorporated in such a diversity of "experiments" as mental tests, projective tests, constancy experiments, absolute and difference threshold experiments, scaling experiments—indeed, all those experiments where the main effects are processes which do not change much with time or experience. There is no feedback in this trial paradigm, the *S* has no knowledge of results within the trial.

The **IRS** paradigm is one learning and concept formation trial used to study those processes which do change with feedback or experience, where we expect learning to occur. In this trial the organism makes an **R** before there is a significant **S** change and then gets feedback in the form of "a reinforcing stimulus" (discovers whether or not he was right, is rewarded or not, punished or not, etc.). This trial is the Humphreys' paradigm trial (5) used by Bush and Mosteller (2) and is the simplest trial type in which learning is a main effect.

The **SIR** trial is very much like the **ISR** trial; in those cases where **S** does not cease before **I** is presented, it may yield identical results in **R**. In the case where there is no overlap in time between S and I, however, the differences in R observed between these two types of trials may be instructive. These differences are attributable to characteristics of set or attitude, "tuning the organism," etc. Experiments where I's both precede and follow S's have been used to demonstrate the existence of "central perceptual selective processes" as distinguished from "response" tendencies and are important in studying search behavior. In a "perceptual selection" experiment by M. Lawrence[3] (9) S's were presented, each of which contained a lot of various numbers, letters, and symbols. The S was a tachistoscopic exposure of a projected slide. In some trials I's to "Look for numbers (letters or symbols)" preceded the brief presentation of the slide, and in some trials the I's "Write down all the numbers you saw" followed the S. Since *S*s were better at selecting the correct elements in the ISR condition than they were in the SIR condition, and since the presentation of the S's was too brief to allow eye movements (search behavior), the difference in results could be attributable to a difference in a central perceptual (selective) readiness process, or to a rapidly declining perceptual "trace." Whether the ISR-SIR differences are due to forgetting or to the presence of the "state of readiness" could not be distinguished by this simple design.

---

[2] These "inconsistent" trial types are: **SRI**, **RSI**, and **RIS**. **SRI** is an "unconditioned response" trial, a simple **S-R** trial with a useless **I** at the end. The **RSI** trial and the **RIS** trial would be "operant" trials also with useless **I**'s. If we stretch the notion of **I** to include those which "act like" reinforcements (in the model, all reinforcements are considered as **S**'s whether or not they are given verbally), we are tempted to make something out of these trials. However, according to the model, they are not trials because the **I**'s are incapable of altering the **R**'s to which they refer (perhaps the **I**'s "remember it" or "forget it" would be an exception, since they refer backwards) and are, therefore, either not **I**'s as defined by the model or are not a part of the trial to which they are supposed to belong. The model places this restriction on **I**'s and only on **I**'s: **S**'s and **R**'s may come anywhere in a trial, but, if there is an **I** (and there must be at least one per trial), it must precede all **R**'s in that trial.

THE IDEAL TRIALS

The SIR-ISR difference brings up an important point about the *ideal* trial for any particular experiment. The ideal trial is a kind of a *statement* of an experiment, at the first level, in the shortest possible form. Theoretically, any experiment should consist of an ideal trial and all the controls which are necessary. The necessary controls are completely derivable from the ideal statement by considering all possible sequences within the ideal trial which obey the rule: at least one of each element must be present in a sequence before it can be considered a trial.[4]

Thus the ideal trial for the M. Lawrence-type experiment for studying the effect of *change in position of* I's *relative to* S is ISIR. For studying the *position of the* S's *relative to* I it would be SISR. From either of these two ideal trials we may derive the two controls of SIR and ISR.

The **ISIR** *trial.* In actual practice, this trial would at times (probably half the time) demand *S*'s switching his set before making a response: "Look for numbers"—exposure of S—"Write down letters"—R. Some of the effects might wash out under these conditions, since *S* would learn not to trust *E*. In any case, before the paradigm has been run in its ideal form, we remain ignorant of much concerning this process.

The **SISR** *trial.* The fact that this ideal trial is possible shows that there are really two alternative theoretical sets contained in the I's in this experiment, one referring forward and one referring backward to the S. As long as the trials SIR and ISR are run separately, this fact is obscured. For the SISR trial to make sense, I must explicitly state to which S the R is supposed to refer. In practice, this trial would sound like: (*a*) exposure of S—"Write down the numbers you have seen (ignoring the next S)"—second exposure—R; or (*b*) exposure of S—

---

[3] Not to be confused with an experiment by Douglas Lawrence and George Coles (8), who used the same trial types but employed a *recognition* procedure, and hence a very simple stimulus, and found no differences between ISR and SIR.

[4] Where differences between events of the same class are not considered, $C = (2^s - 1)(2^i - 1)(2^r - 1) - 1$, where C = number of control trials and *s*, *i*, and *r* are the number of S's, I's, and R's in the ideal trial.

"Write down the numbers you are about to see (ignoring the last S)"—second exposure—R.

The analysis of an experiment, in this way, forces us to discover many of it potentialities and limitations. The "either-or" characteristic of this paradigm, as we have developed it, is evident. That is, the R is restricted at any time to one and only one part of one and only one S, which must be considered as a limitation. It is a study of the ability of *S* to follow singular and exclusive directions, which come at one of two positions in the trial and which instruct him to attend to one of several aspects of one of two S's occurring at different positions. In the form deduced from the model it is a much more complete and exhaustive experiment and tells us much more about the *process* of selective set.

The complete experiments of two ideal trials and two control trials can answer the following questions, among others:

1. If *S*s are better on ISR than on SIR, then there exists a central selective state of readiness or a forgetting process or both. (These cannot be distinguished by studying just these two trials separately.)

2. If, in ISIR, *S*s are better when the first I is the same in content as the second I (e.g., both are "numbers") than when the I's are different, this is not due to a declining trace (although, of course, there may be a declining trace which would best be studied by varying time as a parameter). It may be due to the fact that the same I is being repeated and thus remembered better. This can be checked by comparison with ISR where there is no repetition.

3. If *S*s are no better, in ISIR, when the I's are the same than when they are different, this may be due to one of two causes. First, effects may wash out due to the inclusion of multiple I's (*S*s may not pay attention to the first I after a while); this can be ascertained by comparison with SIR and ISR. Second, if there is no perceptual state of readiness operating, the main effects are due to a declining trace.

4. The SISR trial enables us to see whether it is where the I occurs or in which direction it *refers* which is critical. When I refers to the first S, results should be identical with SIR. When I refers to the second S, results should be identical with ISR. Any departures which obtain are due to the presence of the "uninstructed" S.

MULTIPLE EVENT TRIALS

We will now turn to some trials which can be generated by the introduction of more than one of each of the elements S, I, and R.

An example of a multiple R trial. A recent study by Eriksen (3) is a good example of the appropriate use of a multiple R trial. *S*s were presented with a size-judgment task under the method of absolute judgment. Instead of using only one R or judgment to each S presentation, *E* varied the number and nature of the judgmental R's. *S*s were required first to make a verbal absolute judgment, then to move a lever to a distance proportionate to the size of the presented S, and then to make a "second guess" absolute judgment, verbally. The experiment was designed to find out whether there was an independent relation between each of the several R systems and the S, or whether all of the R's would function together. It might be that the first verbal judgment would serve to define the perception for the *S*, the lever R and the second verbal judgment being mere reflections of the first judgment without any independent relation to the S. Phrased statistically, the question becomes one of determining whether there exist significant partial correlations over trials between each of the R systems and the S when the remaining R systems are held constant. The results showed that the first and second verbal R's and the movement R each had a significant correlation with the S's when either or both of the other R systems were partialed out and that there was some independent error in each of the R's, since all of the R systems together gave a better picture of the S than any two. Considering that the second verbal judgment was the same as the first, except for the fact that it came after the first and after the lever-moving judgment, these results are, in a way, surprising. Eriksen concludes:

> One factor that might constitute at least part of this noncorrelated error is a difference in judgmental task that confronts the *O* at the time of the second verbal judgment. For the first judgment *O* must attempt to deal with a continuum of size and to place the stimulus in one of 11 categories along this continuum. However, at the end of the second judgment *O* is faced with essentially a two-category judgment. He has already located the stimulus along the size dimension and his judgment now is in terms of whether the stimulus is apt to be larger or smaller than his first judgment. This difference in judgmental task may be one of the factors subsumed under the noncorrelated error term (3, p. 358).

Eriksen is tentatively proposing the possibility of the discovery of a very interesting effect: the way in which a complex R is set up and distributed over time is a condition of the accuracy. We might interpret Eriksen's suggestion as that of the presence of something like a "scanning mechanism." The *S* does better if the R is distributed in such a way that he can narrow his search in a systematic fashion—first a gross R, then a finer adjustment, etc.

The event-arrangement model can be used to generate a set of controls for testing the existence of a "contingent-R" mechanism suggested by Eriksen. The basic experimental trial is a double R trial much like Eriksen's, except that we have eliminated the second or "lever" R and replaced it with "intervening activity." We also force the *S* to do what Eriksen suspects he is doing on his second R, that is, to "correct" his first R. The *S* always makes his second R using his first R as a baseline.

The R's in this experiment differ with regard to the number of alternatives potential in each. Furthermore, any R may be present or absent at two places relative to the intervening activity. We are thus free to combine our two R's into a single R (equal, in this case, to the product of the alternatives in the separate R's), and to run single and double R trials for each of the different R's.

It is believed that this paradigm contains a set of controls to account for possibilities not inherent in Eriksen's design. A plan for the experiment is given in Table 1.

Not all of the controls are absolutely essential to show the existence of "something interesting," but, if the results of this kind of experiment produced any

TABLE 1

DESIGN FOR A PERCEPTION EXPERIMENT *

	1st Time Period	2nd Time Period	3rd Time Period	4th Time Period
1. Ideal trial type	S(X)	R(A)	r	R(B)
2. Control trial type	S(X)	O	r	R(AB)
3. Control trial type	S(X)	R(A)	r	O
4. Control trial type	S(X)	O	r	R(B)
5. Control trial type	S(X)	R(A)	r	R(A)
6. Control trial type	S(X)	R(B)	r	R(B)
7. Control trial type	S(X)	$R(A^2)$	r	O
8. Control trial type	S(X)	O	r	$R(B^2)$
And the transpose of each of the above except 5 and 6	S(X)	R(B)	r	R(A)

* S(X) = a stimulus presentation of one of X possibilities where X is, for example, $\gg A$; R(A) = a response of one of A possibilities; R(B) = a response of one of B possibilities where B is significantly $<A$; R(AB) = a response of one of $A \times B$ possibilities; r = intervening activity; O = no activity.

consistent hierarchy and equivalence of the trial types in terms of subjective accuracy, it would enable us to go a long way in specifying the type of mechanism involved in the handling of incoming information. It might turn out, for example, that any split-up of the R is better than no split-up at all; on the other hand, it is possible that only a split-up in which there are significantly fewer (or perhaps even more) categories in the second R than in the first R will yield a higher correlation with the S. We already know that *S*s do better when they have more R-alternatives in which to express themselves. If no other substantial difference is obtained between these trial types in the correlation with the R, then we can propose that, in the particular situation investigated, this multiple IR trial can be reduced to a single IR trial, and many problems, including questions of the appropriate definition of S, I, and R, will be greatly simplified, if not solved.

On a different but related topic, Postman has said:

> . . . it appears that an inference concerning perceptual discrimination can be made only if it is based on *at least* two classes of responses. Thus indices of discrimination based on the analysis of subjects' reproductions should be validated by measures derived from other procedures such as the method of recognition. Similarly, thresholds determined from distribution of verbal responses may be supplemented by measures based on matching responses, and so on. Holding the manipulations of the stimulus and the tuning of the organism constant, we measure variations in at least two classes of restricted responses. To the extent that the two (or more) methods of measurement lead to consistent conclusions concerning the subject's discrimination we have identified a *perceptual response disposition*. The larger the number of response classes which we use in our tests of discrimination, the more secure are we in our inference of perceptual response dispositions. It is not to be expected that two or more classes of responses governed by the same disposition will yield identical indices of discriminations, such as thresholds or constant errors. If the indices differ, the analytic problem is to divide the observed systematic variance in responses into two major components: that due to the common perceptual disposition and that contributed by the availability or probability of the specific responses used (**10**, p. 65).

The multiple IR trial looks promising as an instrument for getting the kind of data which Postman feels is necessary to define a perceptual response disposition.

The $I_1R_1I_2R_2 \ldots I_nR_nS$ *trial.* Eriksen's experiment and Postman's statement suggest a wide application for that type of trial where a number of R's, each with its own I, follow the S. Basically, this trial is still a perceptual trial, since R's, even though there are a great many of them, are still made in the presence of (or following presentation of) the S. Two further possibilities immediately arise. First, would there be any advantage to the multiple IR series in the learning trial? A trial might then consist of having the *S* predict, in a variety of modes and/or the same mode distributed in various ways, what he expects the S to be. An $I_1R_1I_2R_2 \ldots I_nR_nS$ trial would allow us to study learning under those conditions where there may be a lot of R error which does not affect the learning process as far as the *S* is concerned (nonperceptual error), but which otherwise would get in the way of our understanding of the learning process itself. The author wishes to make this suggestion very tentatively, however, since a number of problems arise which are too difficult for consideration here. Instead we shall move on to the second possibility, a multiple IR trial which combines the potential for learning and perception, or discrimination, in the same trial.

*The **IRSR** trial.* The IRSR trial is one in which the *S* is asked to make both a predictive or learning R and an antecedent, perceptual or discrimination R to the same S. Where there is S-R error in both of these systems, the trial paradigm is not trivial. From the point of view of learning, for example, we can ask the question: What happens to learning when the reinforcement is unclear (when there is error in the discrimination of the reinforcing S)? The two- or three-choice learning paradigm (2), in which the *S* is asked to predict whether the S is going to be, say, red or black, can now be extended to those situations where the *S* is also unclear as to the outcome of his prediction, but where the *E* has an S-R measure of this lack of clarity. We tend to think of reinforcing S's as varying in strength, but they also vary in discriminability, and this is an important characteristic to be studied.

The two control trials which can be derived from the IRSR trial are IRS and ISR. A simple example of the use of these three trials to study the relationship between perception and learning might be the following one, in which only two values of S are studied at one time. Let us say that, in any trial, S may have one of two values; that is, it may be big or little, the *S* being asked to state (or predict, depending on the trial used) which of the two values has been (or will be) shown to him. Throughout the experiment one value of S occurs at random more frequently than the other from trial to trial, so that there will be something for the *S* to learn.

First the ISR trial is studied, and a psychometric function can be derived where the difference between the two values of S is the variable. We will then know what percentage of the time the *S* can be expected to distinguish between the two S's in an "absolute judgment" sense, given the physical values of the two S's. We then run the IRS trial, asking the *S* to predict which of the two values *will* occur (big or little), and we do this at selected points on the psychometric function—for example, where the *S* can distinguish 90%, 50%, and 10%. Of course what the *S* can be expected to "learn" is that one of the values of S is more frequent. We can then run the combined IRSR trial to see if there is any difference when it is run as a whole.

Such an experiment can be expected to yield important consequences for studying the relationship between perception and learning. Can learning occur when discrimination is very low, if there are enough trials? Does the nature of the process shift as the reward becomes unclear? Must an event be highly distinguishable for it to act as a reinforcement?

From the point of view of perception, the IRSR trial seems to fit current perception theory much better than the traditional ISR trial. Current perception theory stresses such factors as set, directive state, assumptions, expectancies, determining tendencies, etc., all of which are characteristic of the organism before stimulation. The traditional paradigm allows us to vary these "set" factors only as parameters which can be changed from experiment to experiment, usually through I's. The IRSR trial gives us an R measure (the first R) which corresponds very closely to set, and an opportunity to study changes in set which may occur from trial to trial. Recently there have been suggestions in the perception literature (1, 4) that "responses in the absence of stimuli" will decide nativism-empiricism arguments.

Two interesting correlations may evidence themselves in data collected with the IRSR trial which cannot be obtained with simpler perceptual or learning trials. The first of these is the correlation between the two R's within the trial. The question can be asked: "How much does an individual's directive state, set, or predictive R tend to determine his perceptual R in the case where the other correlations between the two R's and the S are zero, or partialed out?" This is the familiar question of the influence of set, which bears a great deal of similarity to the problem of projection. It is also the same as the general question of how much perception is determined by what the organism brings to the situation, as opposed to critical S factors. It is traditional in psychological experiments to study the influence of set and motivation factors by varying overall I's and by varying differences between individuals by selection procedures. It is conceivable that changes in set occur from trial to trial due to the experiences of the *S*, as well as from experiment to experiment due to alterations in I's and the like. An R measure of set is necessary in order to discover these phenomena if they exist.

*The **ISRS** trial.* The ISRS trial has the same control trials (ISR, IRS) as the IRSR trial and therefore can be considered both from a perceptual standpoint and from a learning standpoint; either as a basically perceptual trial with an added S at the end or as a learning trial with a preceding conditional S. Considered as a learning trial, it is the paradigm of the paired associate or Thorndikian (13) experiment; as a psychophysical trial, it provides the *S* with "knowledge-of-results." Those cases where there are large and consistent errors in the beginning of a run—where knowledge of results in the form of an appended S makes a difference, in the sense that the *S*'s R's become more accurate over trials—are properly considered as studies of learnable skills. In many cases, notably in psychophysics, knowledge of results does not make much difference. Here we tend to feel that we are studying sensation or a perceptual process with little consistent error, in which learning either plays no part or will play no further part, the process having reached a steady state.

Some very interesting perception experiments have been done using this sort of knowledge-of-results trial. The Ames Demonstrations (6) are presented in such a manner that the *S* is first stimulated by a retinal image leading to a perceptual R, which R, although it has a strong feeling of surety attached to it, turns out upon further experience by the *S* to have very little correlation with the physical S. *S*s register a high degree of surprise under these circumstances, and this experiment is very upsetting for some of them. With a trial such as this, one would expect that learning might take place with repeated presentations. If learning can be shown to take place under these circumstances, it is very important for perceptual theory, since what we are dealing with here is a very stable percept of which the *S* is extremely sure and which subjectively feels as though it is highly dependent upon the physical S. If it can be shown that through repeated presentations of this trial a change can occur in that percept, holding other things constant, then questions regarding the extreme adaptability of the perceptual process are automatically raised. Kilpatrick (6) has attempted to show that such learning does indeed take place. His *S*s were presented with a room which looked normal, but which was actually distorted in such a way that the retinal image was almost identical to that produced by a normal room. The R of the *S* was a motor one in which he poked around the room with a stick or tossed balls at spots of light projected on a wall of the room. These actions led to a feedback of a sort very unexpected and discomforting to the *S*. In observing changes in the *S*'s perception of the shape of the room, Kilpatrick concludes that there are two processes: first, a change in the relative weights given to the cues, and second, a complete reinterpretation and reorganization of the cues themselves, so that, when now presented with a similar normal room, the *S*s will see it as distorted. The ISRS trial is an important tool in the study of the relationship between learning and perception, since, like the IRSR trial, it provides opportunities to study change in behavior over trials (learning) in what sometimes is considered an unchanging (perceptual) process.

SUMMARY

A model involving the combination and permutation of the occurrence of events within the experimental trial was suggested which enables one to generate a variety of trial types. The events which make up a trial were proposed as stimuli, instructions, and responses. The definitions of these events were adequate enough to permit the making of certain restrictions on the occurrence of events so that a number of trials, some the same as and some different from those already in use, could be described and their potentialities discussed. The values of the study of trial design lie in its use as an aid in the conceptualization of experiments, particularly in the generation of new experiments and in the providing of a framework within which we can fit experimental findings.

REFERENCES

1. BOLLES, R. C., & BAILEY, D. E. Importance of object recognition in size constancy. *J. exp. Psychol.*, 1956, **51**, 222–225.
2. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.
3. ERIKSEN, C. W. Prediction from and interaction among multiple concurrent discriminative responses. *J. exp. Psychol.*, 1957, **53**, 353–359.
4. HOCHBERG, J. Perception: Toward the recovery of a definition. *Psychol. Rev.*, 1956, **63**, 400–404.
5. HUMPHREYS, L. G. Acquisitions and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
6. KILPATRICK, F. P. Two processes in perceptual learning. *J. exp. Psychol.*, 1954, **47**, 362–370.
7. KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt Brace, 1935.
8. LAWRENCE, D. H., & COLES, G. R. Accuracy of recognition with alternatives before and after the stimulus. *J. exp. Psychol.*, 1954, **47**, 208–214.
9. LAWRENCE, M. *Studies in human behavior.* Princeton: Princeton Univer. Press, 1949.
10. POSTMAN, L. J. *Current theory and research in motivation, a symposium.* Lincoln: Univer. Nebraska Press, 1953.
11. SLACK, C. W. Feedback theory and the reflex arc concept. *Psychol. Rev.*, 1955, **62**, 263–267.
12. SKINNER, B. F. *The behavior of organisms: An experimental analysis.* New York: Appleton-Century-Crofts, 1938.
13. THORNDIKE, E. L. A proof of the law of effect. *Science*, 1933, **77**, 173–175.

(Received June 17, 1957)

Psychological Review
Vol. 65, No. 2, 1958

ON PSYCHOLOGICAL SIMILARITY

MICHAEL A. WALLACH

Harvard University[1]

When Hume spoke of composite ideas, such as *table* or *chair,* as being compounded from simple ideas, such as *brown* or *hard,* he found himself facing the question of how one simple idea could resemble a second one more than a third one. He could handle the question of similarity between composite ideas easily enough; the degree of similarity of two composite ideas depended on the number of simple ideas they had in common. But what about similarity between simple ideas themselves? What does it mean to say blue is more similar to green than to red? Hume could not answer this beyond suggesting that similarity between simple ideas is somehow given to us directly—a capitulation that could never satisfy a British empiricist (1; 11, pp. 585–689).

Let us take Hume's predicament as our starting point. Whence does his problem arise? Its origin lies in two omissions: a failure to note the full range of attributes in terms of which objects and events can be grouped and a failure to separate the question of *potential* similarity from that of *psychological* similarity. Consider the first matter. What are the attributes in terms of which objects and events can be grouped or separated? Hume limited himself here—for a reason noted below—to considering simple sensory attributes displayed by the object itself—for example, the brownness or hardness of a table. But a table can be defined in terms of many more attributes than these. To mention but a few possibilities, it can be defined in terms of use (the things done upon it), of location (the places where it is found), or of construction (how it is built). Such attributes as these are no less reducible to a sensory basis than the ones mentioned by Hume, and they suggest a possible answer to Hume's question. We may say green seems more similar to blue than to red because there indeed *are* more attribute-values that green and blue but not red have in common, while there are fewer that green and red but not blue have in common. Hume may, that is to say, be able to retain his common attribute view of similarity by accepting a wider definition of the term "attribute." Thus, for instance, green and blue—but not red—have in common the attribute-value of being colors of grass in the country, colors of water and hence cool colors, summer landscape colors, and so on.

But we have not fully answered the question yet. Some people may *not* say that green and blue are more similar than green and red. The attributive basis for a judgment of similarity may be present and yet the judgment not be made; and, on the other hand, an attributive basis for such a judgment may be lacking and yet the items may be judged similar. There would seem to be a difference, then, between potential similarity and psychological similarity. Part of Hume's difficulty seems to arise from equating the two. This leads him to fill the perceiver's head with similacral (i.e., photographic) representations of environmental attribute-values, and hence to ignore the possibility that the observer may perform a selecting and ignoring function with regard to attribute-values which are environmentally available. Indeed, it is this view that leads him to the first omission we noted—his failure to specify the full range of environmental attributes. For insofar as one assumes the organism to register in similacral fashion what the environment offers, one must conceive of that environment as relatively meager in its offerings, or else the organism would be swamped with information. If, however, one conceives of the organism as selecting some attributes by which to compare items and ignoring other attributes, then it becomes reasonable to suggest a difference between potential and psychological similarity. Potential similarity may well be most usefully measured in terms of the number of common environmental features that two objects or events are found to display—using the broad definition of attribute noted above, and determining the commonalities by reliable means.[2]

---

[1] The writer is greatly indebted to Jerome S. Bruner for invaluable encouragement and aid in the preparation of this paper. Thanks are also due Albert J. Caron and George Mandler for critical readings of the manuscript. Preliminary research for this paper was done while the writer held a Westengard Travelling Fellowship from Harvard University to the Department of Psychology, University of Cambridge, England.

This definition has its basis in the old view of similarity as common stimulus elements, to be sure—but it is very different from that view, since the common-stimulus-elements approach used the narrow definition of attribute for which we criticized Hume, and that view was thought to apply to psychological similarity, which we shall see is often not the case. Establishment of the degree of potential similarity between two things provides us with an important baseline: It lets us validate or contravalidate individual reports of similarity. A person may, for instance, report that two problems are similar in the sense that the same solution is thought to be applicable to both; to ask whether this feeling of similarity is valid is to ask whether the same solution is in fact applicable to both problems; are they potentially similar in the sense of having this property in common? Or a person may report that two objects are similar in the sense that both are thought usable for driving a nail into a board. To validate or contravalidate this impression of similarity, we determine whether both objects can in fact be used this way.

Broadening the definition of "attribute" and recognizing that potential and psychological similarity are distinct but related issues, suggest that the question of psychological similarity may lie at the basis of diverse kinds of psychological research on cognition that have hitherto usually been considered apart from each other—ranging from studies on stimulus generalization to studies on learning and thinking. We submit it may be fruitful to inquire how some of this research bears on the issues we have raised, and this for two reasons. First of all, we have not yet examined the possible ways in which psychological similarity might be defined, although we have suggested what seems to us a tenable definition for potential similarity. Insofar as this research really bears on psychological similarity, it may help us weigh alternative definitions of the term. Secondly, such exploration of definitions may well have experimental implications for further research in these diverse areas.

SOME DEFINITIONS OF PSYCHOLOGICAL SIMILARITY

One may suggest at least four ways of defining psychological similarity. The first derives from Hume's approach.

---

[2] The problem of just what means to use and how, has been treated at length by Brunswik (3) in his discussions of ecological validity. The means are, basically, responses of a sample of judges—and they may be responses as varied as reading a light meter or judging a painting's colors to be warm or cool. The validation, then, is consensual in nature.

Hume, in equating potential and psychological similarity, supposes that if common properties are present in the environment, they will be perceived by the person. Hence, impressions of similarity can be directly controlled by varying the environment's potential similarity in the sense of the common values of attributes it displays. The selecting and ignoring capacities of the organism are neglected, and psychological similarity is defined in terms of *common environmental properties*. An alternative approach is to recognize the organism's selective functions, and suggest that recognition of common environmental properties depends upon making a common response to instances that share this attribute-value. The attribute-value may be environmentally present in two items and yet the person may not offer a common response to them; or the attribute-value may be absent and yet a common response be made. There is not necessarily a direct relation between psychological and potential similarity. Psychological similarity here, then, is defined in terms of *common responses*. Yet another possible approach is to suggest there are neural traces laid down along various dimensions when a stimulus impinges, and psychological similarity depends on how far a new stimulus is from the old stimulus on such a dimension—the view of *primary stimulation gradients*.

A fourth view may also be proposed. It is related, to be sure, to the view of common responses, but seeks to separate the process of recognizing similarity from making a common response. Yet this fourth approach is not that of stimulation gradients, for it does not assume that there is necessarily an inverse relationship between degree of recognized similarity and distance of new from old stimuli along some dimension. The position may be put somewhat as follows. Although a common response may be forthcoming to signal the recognition of similarity when two events occur, this response is itself a rather trivial matter (it may be made with the big toe as well as with the vocal cords, and need not be of fixed character), and its being made depends on the prior learning of a rule for categorizing events as equivalent. This rule specifies what property (attribute-value) or properties must be conjointly found in order to assign events to the same class, and also tells us the nature of the contrast being made—the relevant alternative class whose members are to be judged nonexemplars of the class of interest.[3] We construct this contrast class in such ways as the following. (*a*) By implication, from the degree of complexity of the class of interest; since the more complex that class is, the more likely we are to consider small deviations from it as the only relevant alternative class or classes, and to totally neglect large deviations.[4] Thus, for instance, we are likely to contrast cocker spaniels with setters or poodles and not with all other animals, whereas "all other animals" would be a relevant contrast class if we were interested in dogs rather than cocker spaniels. (*b*) By experience with all the instances, positive and negative, to which we are exposed, for these establish the domain of material relevant to us. Construction of a contrast class in turn suggests which properties we may best use for discriminating positive from relevant negative instances. For example, while the attribute of size will not let us discriminate dogs from all other animals, it does suffice to separate setters from cocker spaniels.

The classification rule that we learn, then, has two aspects: it tells us on what basis we are to group two events in one class and also tells us the range of other events with which we are contrasting positive instances of the class of interest—it tells us what negative instances are relevant. The rule serves to delimit a particular domain of events, and to draw a *contrast* between two subsets within that domain: positive and negative instances of the class in question. This rule stipulates that certain properties must be found compresent, i.e., together, in order to place an instance in a given class. Such a rule for class-inclusion becomes harder to attain, the greater the number of properties which the rule requires to be compresent before inclusion in the given class is warranted. This is the case because the larger the number of properties whose compresence is required in this way for class-inclusion, the greater is the number of properties which are found compresent in at least some of the relevant negative instances as well as in all positive instances of the class in question, thus making it more difficult to locate the difference between positive and negative instances.

---

[3] See Bruner, Goodnow, and Austin (2), p. 38; and Kelley (13), pp. 303–305.

[4] The "implication" in question here is not one of logical necessity, but rather a strategy that people tend to use—and indeed a valuable one—for determining the level of classifiable category discriminations to make in particular situations.

In brief, then, this fourth proposal suggests that recognition of similarity depends on applying a rule which leads one to *assign items to a common category*. There is much more to similarity than the response of making this assignment. For prior to that, one must learn the rule which guides the response.[5]

Each of these ways of defining psychological similarity has much to be said in its favor, but no one of them seems able to account for all the evidence. Our aim in the following pages is to suggest some possible limitations which various of these views face, and to sketch the possible fruitfulness of exploring more fully the one of these approaches that defines psychological similarity in terms of assigning events to a common category.

If, as we have suggested, diverse areas of psychological research all concern psychological similarity, one is led to ask whether these areas may be using the same kind of procedure, since a common concern sometimes intimates a common procedure of inquiry. This seems indeed to be the case here, and the procedure in question is that of transfer—the influence of experience in one situation on the handling of a subsequent situation. Insofar as something learned in a first situation will aid performance in a second one, then the subject will show positive transfer if he recognizes the two situations to be similar; while insofar as something learned in that first situation will hinder performance in the second, then he will show negative transfer if he recognizes the two situations as similar. Transferability of learning thus provides a set of operations for determining whether recognition of similarity occurred. As we turn to various experiments now, in further considering the definitions of psychological similarity that we have reviewed, we shall find that all of them in effect use this transferability procedure.

COMMON ENVIRONMENTAL PROPERTIES

Since this is, perhaps, the most familiar and earliest view as to the nature of psychological similarity, we need not dwell on the many kinds of evidence from which it arose—for instance, motor learning studies, of which the work by Lewis, McAllister, and Adams (20,

---

[5] If one prefers to argue that such a rule is simply a complex response, it is incumbent upon one to specify the responses constituting that rule. The suggestion often made at this juncture is that the rule is no more than the verbal responses which express it. But in animal learning studies discussed later—where one cannot, therefore, say the rule is nothing more than verbal responses—we shall note that the common responses which are present do not in fact constitute the basis of the rule that is learned.

21) provides a recent example. Commonality of the environmental properties is established by the identity of measurement readings of the centimeters-grams-seconds variety in the two situations being compared (call them Original Learning, OL, and Transfer Learning, TL). The evidence indicates that frequently, to be sure, transferability of learning, and hence psychological similarity, is related to presence of common environmental properties in these two situations. But there are, on the other hand, cases where presentation of common properties in OL and TL does not result in psychological similarity of the two and hence in greater transfer.

Consider, for instance, a study by Schwarz (27). In the OL task, the subject learned to attach a certain response to a certain stimulus. In the TL task, he had to learn to link a different response to that old stimulus. If transfer here depends on common environmental properties, one would expect associative interference—that is, the old stimulus should tend to elicit the old response; it should be transferred to the TL task despite its inappropriateness there. Actually, such associative interference occurred only if the subject was forced to pay attention to something other than this particular task; if allowed to attend to this task, the old response was not transferred despite identity of stimulus. When the subject was "paying attention" he was grouping environmental events differently than when he was "not paying attention." When not paying attention, stimulus identity was allowed to become the attributive basis for classification; but when paying attention, the temporal difference between "stimulus in the OL task" and "stimulus in the TL task" was effective in directing the subject to assign the OL and TL stimuli to different classes based on this temporal distinction, even though they were physically identical. Assigning them to different classes meant that negative transfer was avoided.

Or consider a study by Ellis (4) in which college students learned finger mazes. Three different mazes were used in all, Maze 3 consisting of Mazes 1 and 2 linked together. One group learned Mazes 1 and 3, another learned Mazes 2 and 3, and a third learned Maze 3 alone. Whereas an identical-stimulus-elements approach would predict marked positive transfer for the first two groups in their learning of Maze 3, in actuality those two groups were not significantly better in their learning of Maze 3 than the group learning that maze *de novo*. Objectively identical stimulus elements did not per se lead to recognition of similarity.

Common environmental attributes, then, do not seem to be an infallible indicator of psychological similarity. Such similarity may, under these conditions, be present for the person—but then again it may not.

COMMON RESPONSES

It was perhaps the Kantian revolution in philosophy that first called into question Hume's conception of the organism as a mirror of environmental stimulation, and suggested instead that the organism is a filterer and arranger of environmental events. The "mentalism" of Kant's approach banned it from having direct repercussions in American psychology until it was realized that considering the organism an active mediator does not require descent into its depths, but rather can be achieved by noting how the organism responds to its environment. One need not go beyond the relation between environmental stimuli and a person's responses. If, according to this view, a person was found to respond the same way to two objectively different situations, then those two situations were psychologically similar for him. And if, on the other hand, a person responded differently despite repetition of the same situation (common environmental properties), then no such psychological similarity obtained. Such an approach constituted a tremendous advance. The Humian assumption of man as capable merely of similacral representations of his environment was overcome, and in its place the capacities of the organism to select, ignore, and emphasize environmental attributes were recognized. The exciting effects of this idea can be seen, for example, in the early work by Shipley (28, 29). In one experiment, linking an eyewink response to both light and shock, where shock elicited finger withdrawal, resulted in light becoming adequate to elicit finger withdrawal too. The common response of winking made to both light and shock indicated their psychological similarity for the subject. Whereas we would say this response was a signal of this similarity, proponents of the common response view would hold that this response itself constitutes that similarity. The implications of this kind of approach are ably developed by Osgood (25), among others, and its applicability to many questions of psychological similarity is clear.

The common-response view leaves us, however, with some uneasiness. For one thing, those who maintain it sometimes seem to rely on the magical efficacy of the term *response* to provide an aura of explanation where none really exists. The premise of the approach is that transfer will occur, and hence two events will seem similar, to the extent to which the organism makes the same response to both. But from here on the going becomes less simple. Paradoxically enough, the insufficiency of this premise seems to be tacitly recognized, for the full view consists of an attempt to defend the premise's lacks. Since the hypothesized common response that supposedly mediates the transfer sometimes cannot be located even though transfer is found, one meets the rationalization that sometimes the common response linked to disparate stimuli is not observable, but in all such cases these unobservable responses are supposed to have developed as fractional anticipations of observable ones. This is an empty-organism theory with a new twist. The organism gets filled eventually, but only with fractions of things that were overt responses once.

It seems relevant to remember that much of the impetus behind this view derives not from considerations of scientific adequacy as such, but rather from the cultural milieu of an earlier behaviorism in which the model of the organism as capable of transformed representations of its environment was still held in disrepute, as "mentalistic." The common response theory hence sought to make this model acceptable to the scientific community by framing it in the language of responses. This end has been accomplished long since, however, so it would perhaps be advisable now to restrict more carefully use of the term "response" to its usual meaning: muscular and glandular reactions. We would note that the restraints imposed by the nature of muscles and glands imply that behavior must be more limited and have fewer dimensions about it than what goes on inside the organism.[6] Why, then, limit on a priori grounds one's conception of what goes on in the organism to fractions of behavior? It does not make constructs more scientific to frame them in the same terms as observables. Rather, the reason we postulate constructs is that the relations among observables are extremely complex, and it often proves necessary to invest our constructs with properties different from those of behavior in order the better to predict that very behavior. The interests of explanation are not served simply by filling the organism with analogues of behavior. Hospers (9) and other philosophers have well pointed out that such reduction to the familiar is not what we really mean by scientific explanation.

---

[6] One way to indicate this difference between behavior and what goes on inside the organism, is in combinatorial terms: there are more elements that can be combined, and hence a greater variety of possible outcomes, at the neuron level, than at the level of muscles and glands.

But were our uneasiness over the common-response approach to stem only from general considerations such as these, it might well be dismissed. Dismissal becomes more difficult, however, when one notes the gradual accumulation of experimental evidence that questions the sufficiency of response identity for mediating transfer of learning. Consider, for instance, a recent study by Bruner, Mandler, O'Dowd, and Wallach.[7] Rats were overtrained to run left-right-left-right at successive choice points of a linear maze, and then were faced with the task of learning to run right-left-right-left in the same maze. The object was to find whether there was positive transfer from learning the first turn sequence to learning its mirror-image reversal.

According to the response-mediation approach, transfer is said to depend on identity of response, and the greatest identity of response between OL and TL here is the single-alternation weaving motion whereby the rat alternates from one side to the other at successive choice points. The faster a rat runs from start box to goal box, the more of a weaving motion he learns. Insofar, then, as transfer depends on the identity of a weaving-alternation response in OL and TL, rats running faster should show more positive transfer than rats running more slowly. The results do not support this prediction. There was no correlation between speed of running during OL and degree of positive transfer on TL, within any of the groups run. Furthermore, considering differences in running speed between groups—differences that were created by varying hunger conditions—groups whose drive conditions resulted in their running slowly during either OL or TL or both showed positive transfer, while the group whose drive conditions resulted in its running fast during both OL and TL showed no transfer. Apparently, therefore, something other than a common response of weaving single alternation causes the transfer in the present experiment.

Or take North's finding (23, 24) that the performance of rats on discrimination reversals was no better when the rat was made to correct overtly his mistakes than when mistakes were left uncorrected. It was expected that reversal learning would be better in the former condition, since it permitted a correct response to occur after a wrong choice, and this correct response was supposed to become associated to the present maze stimuli, and hence be evoked by these stimuli on the next trial, thereby mediating a full-fledged overt correct response. Identity of response thus was expected to mediate reversal learning. Much positive transfer in the learning of successive reversals occurred, but it was not due to response mediation. It occurred, rather, only when there was a sufficient amount of overtraining on each reversal—a result also obtained in the study on single alternation cited above. Apparently, for the animal to classify initial and later situations in terms of the more abstract property which they in fact have in common, he must be given enough practice on the former situation so that its more concrete properties no longer require all his attention. The abstract property in question was single alternation in the case of the first study noted; reversal of position response on encountering an error, in the case of the second study. The concrete properties involved were particular sorts of muscular kinesthesis that occur in relation to particular olfactory and visual cues from various parts of the maze.

---

[7] Bruner, J. S., Mandler, Jean M., O'Dowd, D., & Wallach, M. A. "The role of overlearning and drive level in reversal learning." To be published.

To be sure, insofar as one broadens the definition of "response" to include anything going on inside the organism, as some theorists such as Osgood (25) and Mandler (22) advise, evidence of the above sorts does not run counter to the common-response view. But broadening the definition in this way would seem more to cloud than to clarify the issues. There are grounds, then, for suggesting that presence or absence of common responses is not an infallible indication of presence or absence of psychological similarity, although there are certainly cases where such responses do serve in this manner.

PRIMARY STIMULATION GRADIENTS

Many studies have been done in support of a view of psychological similarity which may be stated somewhat as follows. When a stimulus impinges on a receptor and becomes associated to a response, a neural gradient is laid down such that other stimuli will also elicit this response but to lesser degree, response elicitability decreasing with increasing distance of the new stimulus from the training stimulus along some dimension of stimulus differences—for instance, intensity or pitch of tones. The dimension of primary stimulus generalization need not be innate, as was assumed by those proponents of this approach who were most directly in the Pavlovian tradition, like Hull (10); but rather may be the product of learned differentiation (i.e., discrimination training)—a coming into closer contact with stimulus differences—as in the manner suggested by Gibson and Gibson (5). But however the dimensions are thought to arise, the position in question assumes the shape of the gradient laid down upon them to fall off on either side of the point of stimulation along these dimensions; and, if training and test stimuli are presented at discriminably separate points along a dimension which the individual has learned to discriminate, or can innately discriminate, then generalization is expected to occur along that dimension.

These ideas of a normal distribution of intensity of excitation about a point of stimulation, and the automatic fashion in which generalization is expected to occur between discriminable points on any discriminable dimension when these points are stimulated in succession, were derived from an earlier era of physiological research in which spatial representations of sensory qualities in the cortex were a prime concern—we may see it in such apparently opposed currents of thought as Pavlov (26) on the one hand, and Köhler (19) on the other. And some work, to be sure, does seem best interpretable in terms of such ideas (e.g., Guttman and Kalish, 8). But there appear to be other studies which, although ostensibly dealing with the same processes, may perhaps be better interpreted in different terms. In a number of experiments on "stimulus generalization" in humans, psychological similarity may be mediated not by that process but rather by the class to which the subject assigns various inputs. The magnitude of the subject's response to a test stimulus can depend, one may suggest, on whether or not he places both training and test stimuli in a common class. There may well be individual differences as to whether these two stimuli are placed in the same category or not; and in addition, experimental conditions may be set up in such a manner as to induce or discourage the placement of those two stimuli in the same class without varying these stimulus presentations themselves. If such diffuse factors as these were to cause differences in magnitude of response to the test stimulus—despite the facts that the dimension along which the training and test stimuli fall and the differences along that dimension are easily discriminable by the subject, and the training and test stimuli presented are the same for all these subjects—then the sufficiency of a stimulus-generalization explanation for such data would be called into question. And, in fact, there do seem to be cases where differences of these kinds occur.

We may note, for example, a study by Wickens, Schroder, and Snide (32), using the GSR. Adult subjects in one group received shock-reinforced presentations of a particular tone, and then were given extinction trials with one of three tones at varying frequency distances from the training tone. Subjects in another group underwent the same regimen, except they heard a nonreinforced click interspersed among the reinforced tone presentations during training. The group not hearing clicks—qua group—yielded a gradient of decreasing GSR magnitude as distance between training and test stimuli increased; but the height and shape of the gradient differed on different extinction trials, and only by averaging the results of large numbers of subjects (24 in each extinction subgroup) could one tease out even these not very great consistencies. The group hearing clicks showed no clear gradients at all, even as a group.

Why the different results from these two conditions? Recall what we noted earlier concerning the two aspects of rules for classing events as equivalent: that they specify the range of negative instances we are seeking to discriminate, as well as the nature of positive exemplars of the class of interest. We would suggest that presence versus absence of clicks during training resulted in the learning of different rules for the classification of test stimuli during extinction. The group hearing clicks used the property "tones" as definitive for class-inclusion, and excluded clicks. All test tones were assigned to one class. Because of what they experienced during the training series, the domain of instances for this group consisted of "tones" and "clicks." The group not hearing clicks, on the other hand, used the properties "tones, and within such-and-such range of pitch" as definitive for class-inclusion, and excluded tones of other pitch—these latter being the relevant negative exemplars for them. Only some test tones met the specifications for this class. Because, then, this group experienced only the repeated tone during training—i.e., heard no clicks—this group's domain of instances consisted of "tones of various pitches." Hence, we would expect the GSR to the test tones far from the training tone to be much less than that to the test tone closest to the training tone for the group not hearing clicks—and this was in fact the case. These subjects assigned only the nearest test tones to the same class as the training tone, further-away test tones being excluded. But we would expect no clear trends of differences in GSR among the test tones for the group hearing clicks—and this also was the case. These subjects assigned all test tones to the same class as the training tone, clicks being the items which were excluded from the class. The difference between the click and no-click conditions, then, seems to be that the subject was induced to classify the tones differently in each case.

It may be that many other "generalization" experiments also involve a categorization process like that we postulate for the "no-click" group in the above study: Test stimuli close to the training stimulus are placed in one class with the latter, those further from the training stimulus being excluded. The widths of the class in question vary across individuals, and hence one finds a smooth normal-curve type of gradient only when one averages all subjects together. But there are further individual differences in classification, and, what is more, classifications may change as a test series proceeds. Hence one finds, for instance, nonconsistent results over extinction trials taking groups of subjects as wholes—as in a study by Grant and Schiller (7). That these effects are due to categorization differences seems a reasonable hypothesis when one considers work like that by Wickens et al., where, on our interpretation, the probabilities of alternative rules for classification were varied, and this was found to cause differences in magnitude of response to a test stimulus.

Although primary stimulation gradients seem able to account for some cases of psychological similarity, then, there are others in which a different process may well be at work.

ASSIGNMENT TO A COMMON CATEGORY

Each of the three definitions of psychological similarity noted above has been found to have only limited applicability. Such is also the case with this fourth definition. Since, however, it is younger than the others, its implications have not yet been thoroughly explored; and since it seems relevant in many cases where the other definitions cannot easily be applied, it seems advisable to examine it further with regard to its range of utility.

According to this fourth definition, when we have attained a rule which bisects a domain of events into a set that exemplifies a certain property or conjunction of properties and a particular contrasting set that does not, then we will transfer behavior from one to another of two events that fall into the former set, but will tend not to transfer behavior from either of them to an event in the latter set. Again, then, the relation has two aspects: It tells us the range of negative instances which are relevant for our purposes of discrimination, as well as the basis on which we are to declare events positive exemplars of the class of interest. In short, recognition of the similarity of two events depends on their being classed as equivalent—as exemplifying the property or conjunction of properties in question, and as standing in contrast to a particular range of nonexemplars.[8] We learn to take one or more particular properties as criterial for sorting a domain of events.

There would seem to be two main determinants of the kinds of classification rules we attain: the nature of the instances presented (i.e., the environmental properties to which we have been exposed); and our predilections for some bases of classification and prejudices against others. Varying either the nature of the instances or our classification preferences hence should influence psychological similarity in predictable ways.[9] Let us see if this approach seems to contribute anything in considering two psychological issues: the first, transposition—a matter around which much empirical study has centered; and the second, distinctive similarity—Köhler's demonstration of the importance of similarity in guiding attempts at more complex human problem-solving.

---

[8] This view is related to those of Goodman (6) and Kelley (13), and was first approached in Klüver's early work (14).

[9] Since the present paper's aim is to compare this fourth approach to psychological similarity with the other three and to provide initial indications of that approach's empirical utility, we cannot discuss this approach in greater detail than is provided in the present paragraphs and in the earlier section where this view was introduced. More detailed discussion of this view in particular will be provided in a subsequent paper.

Transposition

Some years ago, Köhler (**17**, pp. 217–227) and Koffka (**15**) argued that discriminating in terms of such relations as "brighter than" or "larger than" is more primitive, more basic, than discriminating in terms of absolute sensory qualities such as brightnesses and sizes. Spence (**30**) argued just as insistently, however, that responding on the basis of absolute stimulus qualities such as brightness or size is more primitive, and relational responding is a result of the algebraic summation of response tendencies to absolute stimuli. But one may suggest a third alternative. Perhaps higher organisms (and certainly humans) can learn to respond either in terms of relations between stimuli or in terms of absolute stimulus qualities, and which selection they will make depends largely on the conditions instituted concerning classification preferences and the nature of the instances.

Consider, for example, the nature of the instances. We would predict a greater tendency for subjects to respond relationally on a transfer test, when trained on couplets of one large and one small form apiece, and when the particular forms used vary in size (but not shape) across couplets, than if the same amount of training were given on one repeated couplet of a large and small form; in both cases response to the larger form is always rewarded during training. In the former condition, the nature of the instances is such as to permit the subject to rule out the possibility of classifying the data in terms of the absolute stimulus quality of size, and the only class into which all data fit is one defined by the relational property of "larger than"; whereas in the latter condition, the kinds of instances presented permit classification on the basis of absolute size or relative size. We could just as well, of course, arrange our instances so as to rule out the possibility of classifying on the basis of relative size, simply by always rewarding choice of the same particular stimulus form, and having it appear sometimes paired with a larger form, sometimes with a smaller form. In a like manner, whether the subject will classify on an absolute, relational, or even some other basis depends on his biases for or against particular bases of classification, as induced by instructions or other means. Given variation in the nature of the instances and in such biases, it would seem one can create conditions that will induce the subject to classify the training instances, and hence to transfer, on any basis we desire.

Thus, work with children by Jackson and Jerome (**12**), for instance, indicates that how the child categorizes the training instances depends on whether he is asked to decide which stimulus is correct after the presentation of both, or rather is required to indicate, after each stimulus is presented, whether it is the correct or incorrect stimulus. As we would expect from the differing emphases on relational versus absolute classifying implied by these different conditions of instance presentation, relational choices were preponderant when the child gave one over-all response after simultaneous or successive presentation of both stimuli, whereas absolute choices were in the majority when the child had to respond to one stimulus before the second was presented. Requiring the subject to make a separate response to each stimulus favors classifying on an absolute basis, whereas requiring the subject to make a single response to both stimuli as a pair favors classifying on a relational basis.

Stevenson, Iscoe, and McConnell (31) found that college students varied in the basis they used for responding on the first transposition test trial; many, to be sure, responded on a relational basis, but some responded on an absolute basis. Apparently, different subjects received different impressions about what was required of them, and hence classified the data presented to them in different ways. What seem needed in this area, then, are studies analyzing the conditions of instance presentation and biases concerning particular attributes that influence the basis the subject chooses for classifying the training stimuli—conditions which therefore influence the kind of transposition shown.

Distinctive Similarity

Köhler (16) uses the term "distinctive similarity" for denoting that particular items resemble each other more than they resemble the rest of the situation. The nature of this greater or lesser resemblance is itself, however, left unanalyzed, although an earlier experiment by Köhler and von Restorff (18) had explored the effects of such similarity on complex human problem-solving. Two groups were taught a principle for solving an arithmetic problem, and transfer was measured later to the solution of a new arithmetic problem on which that principle—an arithmetic shortcut—could also be used. Between these two occurrences, one group was given arithmetic problems to which, however, this principle was not applicable, whereas the other group was given entirely different tasks to do—the solution of matchstick puzzles. The latter group showed greater transfer of the principle to the new problem than did the former group. Thus similarity of the first to the last problem, in particular, was greater for the latter group than for the former group; in one case it was "distinctive," in the other case it was not. But of what does this psychological similarity consist?

One may suggest that analysis of the classification tasks facing the subjects in Köhler and von Restorff's groups casts light on something which these authors really left shrouded in mystery—the nature of distinctive similarity. After being shown the application of the principle to the initial arithmetic problem, the task of the subject became, essentially, that of attaining the rule specifying what property or properties must be conjointly found in order to assign problems to the class to which this principle is applicable, and specifying the relevant range of negative instances with which the subject must concern himself. The group performing better on the transfer test learned but a simple rule for classification: only one obvious property made an item a positive exemplar of the class of items to which the shortcut principle was applicable—namely, that the item be an arithmetic problem; and the degree of overlap of positive and relevant negative instances in terms of compresence of properties was minimal—no more than that they were all "problems." The group performing less well on the transfer test was given the task of learning a more complex rule for classification; a conjunction of properties was required to make an item a positive exemplar of the class in question: the item had to be an arithmetic problem *and also* had to exhibit certain additional features. For this group there are more properties compresent for positive and relevant negative instances (e.g., that the task be not only a problem but also an *arithmetic* problem is compresent for both) than for the other group. In the first case, the domain of relevant events is "arithmetic and matchstick problems," and the contrast is between these two types. In the second case, the domain is "arithmetic problems," and the contrast is between arithmetic problems with certain features and arithmetic problems with other features. Again, the nature of the instances presented to the subject determines the kind of classification attempted. In short, "distinctive similarity" of two events means that the rule for assigning them to the same class provides for a minimum number of properties to be found compresent in at least some of the relevant negative instances, as well as all the positive instances of that class. The difference between positive and relevant negative exemplars is easier to locate.

Could the rule for the group that does poorly be rendered easier by a change in biases of the subjects toward particular attributes? Let us consider the following variation on the Köhler and von Restorff experiment. Suppose we give another group all mathematics problems and in the same order as the original group, but tell them, after the shortcut has been explained, that not all the subsequent problems are amenable to the shortcut, but that some will be. We have thereby set these subjects to remain vigilant throughout the series for the features of later problems that are the same as those in the initial one—the features that hence permit the principle to be applied—and would expect them to apply the shortcut to the last problem more often than did the group not given this set. Without changing the problems, then, we would expect the percentage of shortcut solutions to increase because of this change in set; these subjects have been more sharply tuned to try to find what features of arithmetic problems permit application of the shortcut. They have been encouraged to try to take a particular conjunction of properties as criterial for sorting events.

If one wishes to study distinctive similarity in the transfer of principles for problem-solving, then, the recommendation would seem to be this. Let us determine the conditions of instance presentation and attribute biases that aid or hinder the arrival of the subjects at rules for classifying tasks in terms of properties which they potentially have in common—conditions which aid or hinder, in other words, a match between psychological similarity and potential similarity.

CONCLUSION

The aim of this discussion has been twofold. First, after pointing out the ubiquitousness of the issue of psychological similarity in various areas of research on cognition, to note the way in which a common procedure, that of the transfer experiment, tends to unite them, and to urge that they be conceptualized in an integrated manner. Second, to consider the values and difficulties of four conceptions of psychological similarity, defined in terms of common environmental properties, common responses, primary stimulation gradients, and assignment to a common category. All four definitions are, to be sure, of limited applicability—but the last of them appears to apply in many instances where the first three do not. Several experimental implications of defining psychological similarity in terms of assignment to a common category were explored, and were taken to indicate that further work with this approach may well prove fruitful.

REFERENCES

1. Aaron, R. I. *The theory of universals.* Oxford: Clarendon Press, 1952.
2. Bruner, J. S., Goodnow, Jacqueline J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.
3. Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

4. Ellis, W. D. Memory for physically identical elements in human maze learning: A transfer problem. *Psychol. Bull.*, 1939, **36**, 545–546.
5. Gibson, J. J., & Gibson, Eleanor J. Perceptual learning: Differentiation or enrichment. *Psychol. Rev.*, 1955, **62**, 32–41.
6. Goodman, N. *Fact, fiction, & forecast.* Cambridge: Harvard Univer. Press, 1955.
7. Grant, D. A., & Schiller, J. J. Generalization of the conditioned galvanic skin response to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.
8. Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.
9. Hospers, J. On explanation. *J. Phil.*, 1946, **43**, 337–356.
10. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
11. Hume, D. An enquiry concerning human understanding. In E. A. Burtt (Ed.), *The English philosophers from Bacon to Mill.* New York: Modern Library, 1939. Pp. 585–689.
12. Jackson, T. A., & Jerome, E. A. Studies in the transposition of learning by children. VI. Simultaneous vs. successive presentation of the stimuli to bright and dull children. *J. exp. Psychol.*, 1943, **33**, 431–439.
13. Kelley, G. A. *The psychology of personal constructs.* Vol. 1. *A theory of personality.* New York: Norton, 1955.
14. Klüver, H. *Behavior mechanisms in monkeys.* Chicago: Univer. Chicago Press, 1933.
15. Koffka, K. *The growth of the mind.* New York: Harcourt, Brace, 1928.
16. Köhler, W. *Dynamics in psychology.* New York: Liveright, 1940.
17. Köhler, W. Simple structural functions in the chimpanzee and in the chicken. In W. D. Ellis (Ed.), *A source book of Gestalt psychology.* New York: Humanities Press, 1950. Pp. 217–227.
18. Köhler, W., & von Restorff, Hedwig. Analyse von Vorgängen im Spurenfeld. II. Zur Theorie der Reproduktion. *Psychol. Forsch.*, 1935, **21**, 56–112.
19. Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
20. Lewis, D., McAllister, Dorothy E., & Adams, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–268.
21. McAllister, Dorothy E., & Lewis, D. Facilitation and interference in performance on the modified Mashburn apparatus: II. The effects of varying the amount of interpolated learning. *J. exp. Psychol.*, 1951, **41**, 356–363.
22. Mandler, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.
23. North, A. J. Improvement in successive discrimination reversals. *J. comp. physiol. Psychol.*, 1950, **43**, 442–460.
24. North, A. J. Performance during an extended series of discrimination reversals. *J. comp. physiol. Psychol.*, 1950, **43**, 461–470.
25. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
26. Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
27. Schwarz, G. Über Rückfälligkeit bei Umgewöhnung. *Psychol. Forsch.*, 1927, **9**, 86–158.
28. Shipley, W. C. An apparent transfer of conditioning. *J. gen. Psychol.*, 1933, **8**, 382–390.
29. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **10**, 337–357.
30. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
31. Stevenson, H. W., Iscoe, I., & McConnell, Claudia. A developmental study of transposition. *J. exp. Psychol.*, 1955, **49**, 278–280.
32. Wickens, D. D., Schroder, H. M., & Snide, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–56.

(Received July 1, 1957)

Psychological Review
Vol. 65, No. 2, 1958

CONFORMITY, STATUS, AND IDIOSYNCRASY CREDIT[1]

E. P. HOLLANDER[2]

Carnegie Institute of Technology

Something of a paradox exists in the prevailing treatments of conformity and status. Students of social psychology are likely to be left with the pat impression that the freely chosen leader conforms to, and perhaps tenaciously upholds, the norms of his group. Yet this kind of leadership is also presented as a status sufficient to provide latitude for directing and altering group norms (11, p. 416). From their recent experimental work in this area, Dittes and Kelley have voiced a doubt that the relationship between conformity and status is ever a simple one (7, p. 106). The evidence favors their assertion.

Although these phenomena may be treated as discrete entities, they both arise from interaction between an individual and a set of relevant other individuals constituting a group. To say that an individual conforms, or that he has status, is not to say that these are independently determined states nor that they are terminal; they have some common origin in a phenomenal relationship which persists over time. Conformity and status may be thought of therefore as mutually dependent, and transitionally effective upon subsequent interactions. With this as a framework, several general conceptions will be expressed here regarding mechanisms which produce these phenomena and govern their relationship to one another.

In a gross way, three classes of variables, or elements, are necessary to this conceptual scheme: characteristics of the individual himself; characteristics of the group with which he interacts; and outcomes of interaction representing a past history which may alter the relationship of the former elements.

Of particular importance as a mediating process is the changing perception brought about in the individual and the group by their interaction; the third element is, in effect, this process. A distinction is required, therefore, between the phenomenal and perceptual features of behavior. An individual's behavior is not only phenomenally present in interaction but is also subject to view and appraisal by the other members of the group. If there are to be consequences involving these others, it is essential that there be a perceptual intake on their part. And so too must the individual perceive a group norm; the fact that it is manifestly there is not enough.

It is worth emphasizing that the focus here is upon how the individual fares in the group rather than upon more global consequences to the group. Two kinds of interlocking mechanisms are of concern: those giving rise to behavior in conformity with group demands, and those giving rise to status. The issues at stake may be put simply as follows: What produces conformity? And what allows for nonconformity?

---

[1] A paper derived from this material was presented at the 1957 International Congress of Psychology in Brussels, Belgium. Thanks are due to the overseers of the Group Psychology Branch of the Office of Naval Research for their encouragement and support of this work under ONR Contract 1849 (00). As usual, the views expressed are those of the author and are not to be construed as necessarily reflecting those of the Navy Department.
No theoretical conception exists in isolation. Grateful acknowledgment is extended to T. M. Newcomb, whose work has been a particular stimulus, and to J. F. Muth and H. A. Simon for their critical review of concepts and presentation. For any inelegancies of formulation or errors of content which appear, the author alone is responsible.

[2] On leave 1957–58 as Fulbright Lecturer in Psychology, Faculty of Letters, Istanbul University, Istanbul, Turkey.

SOME QUESTIONS ON CONFORMITY

Fundamental to these issues is the matter of determining *when* an individual may be said to be conforming. One may note that a twofold assumption underpins the usual view of conformity, i.e., that the individual is aware of the existence of a given group norm, and that his behavior in accordance with this norm is evidence of conformity. It is doubtful that both features of this assumption necessarily hold simultaneously. This being so, difficulties of interpretation will arise. If the individual were to be insensitive to the norm he could hardly be said to be conforming to it, whatever his behavior seemed to betray; correspondingly, a kind of "conformity" might prevail in terms of adherence to an incorrectly perceived norm; and thus, an evident failure to conform might or might not be "nonconformity" depending upon the accuracy of the individual's perception of the norm in the first place.

A related question concerns the individual's motivation. Is there a motive for nonconformity identifiable? Insofar as they are distinguishable, is it necessarily so, after all, that a conflict obtains between the individual's dispositions and the group's demands? Since behavior is taken to be more than a random event, the motivation for instances of conformity or nonconformity should be accountable, once the presence of an adequate recognition of the norm is established.

There remains too the question of who perceives a given behavior to be conforming, i.e., an external observer, a group member, or the actor himself. Employing a fixed-norm baseline for observation, as is often done, serves to obscure differential expectations which render conforming behavior for one individual nonconforming for another—with regard, that is, to others' perceptions *in situ*. Thus, the degree of familiarity with the unique properties of the group context is critical in verifying and understanding conformity.

NORMS, ROLES, AND GROUP EXPECTANCIES

The usual conception of conformity examined here requires some group referent and a standard of behavior abstracted therefrom and defined as a norm. Probably because many studies of groups have involved highly manifest behaviors, norms are conceived to be quite literally evident. On the other hand, in the related concept of role a recognition exists that the behavioral standard may not be manifest, but rather may be an *expectancy*.[3]

Though persisting, the distinction between norms and roles is neither essential nor easy to maintain (cf. 2, 16). Roles are normative in that they involve some implicit shared expectancy among group members; and norms themselves, lacking visibility, may nonetheless dwell in expectancies. It is these expectancies, then, which may be normative, in the sense of typicality. Norms and roles are only distinguishable insofar as norms usually imply expectancies applicable to many persons, while roles are expectancies restrictive to one or a very few individuals in a group.

Objective observers might delimit common expectancies appropriate to group members in general from differential expectancies having reference to particular individuals as such. For the individual in the setting, however, manifest conformity probably comes about without regard to a separate awareness of norms as distinct from roles, but more likely in terms of behaviors which he perceives to be expected of him by relevant others, i.e., "doing the *right* thing."

---

[3] The term "expectancy" refers to another's perception of some object person (cf. 21). What the object person then perceives to be the expectancy is quite important, but its locus is first of all in the "other." Reference is not made, therefore, to the term in Tolman's sense.

In the world of daily interaction, the perception an individual holds of what relevant others expect of him is a singularly important determinant of his social behavior; and the degree to which an individual perceives the group to be rewarding serves to enhance or elaborate the effect produced by his motivation to belong. An alternative sequence may be seen to occur as well: motivation having reference to some fulfillment through the group serves to heighten the individual's perception of its expectancies.[4]

INDIVIDUAL VARIABLES

Granted that conformity derives from certain features of individual perception and motivation, it still remains necessary to identify these features more pointedly. In this formulation, there are four such to be noted: perceptual ability (P_a), representing a general alertness to the social stimulus field; perceptual error (P_e), with particular reference to group expectancies; motivation to gain or sustain social approval (M_a); and, motivation to take part in the focal activities of the group (M_g). Taken together, the latter two variables may be considered as the individual's motivation to belong to the group (M).

The perceptual variables can readily be related to personality typologies. Many of these, e.g., authoritarianism, rigidity, or empathy, appear to lend themselves to a reduction to perceptual function as a core element (cf. 1, 3, 19). Terms like "perceptual rigidity," "perceptual defense," and "social imperceptiveness," often appear as concomitants of these broader characterizations; evidently, this element accounts for certain diversities in behavior which distinguish individuals from one another.

It is useful here, however, to recognize a differential between that which is given and that which is emergent, i.e., perceptual ability and perceptual error, though the interaction of the two is not challenged. The distinction basically is that the former serves as a parameter setting the lower limit on the latter. Thus, the minimum level of an individual's P_e is set by his basic capacity, P_a. This should not be taken as neglect of the potentials of learning, however. The concept of capacity introduced here may be understood to be similar to that of cognitive structures (cf. 15). No assumptions are made about the source of the "capacity"; it is only significant as a feature of the individual which bears upon interaction. It seems reasonable to believe that some individuals have an initial advantage over others as regards accuracy in perceiving group expectancies.

Concerning motivation to belong, mention has already been made that it involves two continua: motivation specific to the activity—or instrumental features—of the group, M_g; and motivation rooted in a generalized need for social approval, M_a. This view cuts across a number of other motivational schema suggested elsewhere (cf. 4, 8, 9, 12, 23), and is intended more as a resolution than a departure. Briefly, these other distinctions appear to involve an "activity focus" and an "other people focus." Activity involves others, of course, but not necessarily to gain their approval. What really seems to matter is the nature of the reward sought.

---

[4] The work on selective perception (e.g., 18) sustains some such formulation, in general. But the linkage between motivation and perception has considerably greater complexity, as Bruner (5) has more recently pointed out.

The approval variable might be viewed as a parameter of personality, but not one so static as to be unaffected by interaction, within certain limits. Since those members having interests which can only be satisfied through participation in group activity do not of necessity have a high need for social approval, and since those cast into groups of little positive activity valence to them may still require approval, it is possible that these variables may be related negatively or positively, depending upon the circumstances considered.

STATUS EMERGENCE

The foregoing points have concentrated on individual characteristics that absorb and deal with features of the social context. Ultimately, these have consequences in behavior, which in its turn has an impact upon the group. It is appropriate now to consider the implications of this process to the emergence of status.

At bottom, status may be taken to be an outcome of the group's differentiated perception of the individual, leading to a set of particularized expectancies regarding his behavior. This occurs as a function of certain of the behaviors or characteristics evidenced by the individual in interaction, which then yield a reconstruction of the group's perception of him. Cast in these terms, status has special value as a kind of middle ground in relating the individual to the group. It exists in the first place as a feature in someone's perceptual field, for without reference to a perceiver status has no intrinsic value or meaning in itself. And, similarly, role cannot be divorced from its perceptual locus; behavior is only appropriate to status insofar as someone perceives it to be so. Perceptual differentiation by the group has consequences, then, in terms of the behaviors it expects the individual to display.[5]

Though not necessarily the case, it is desirable to conceive of status within this framework as having hierarchical properties on some sort of group-acceptance continuum (cf. 7). This is by no means critical as a feature, but is of heuristic value. Still further, it is convenient to represent status as permitting greater latitude in the manifestation of behaviors which would be seen to be nonconformist for the other members of the group; we refer here to common expectancies, a term introduced earlier. The implications of this aspect of status are of especial relevance to what follows.

IDIOSYNCRASY CREDIT

Status will hereafter be considered to be an outcome of interaction referred to as "idiosyncrasy credit" (*C*). This represents an accumulation of positively-disposed impressions residing in the perceptions of relevant others; it is defined operationally in terms of the degree to which an individual may deviate from the common expectancies of the group. In this view, each individual within a group—disregarding size and function, for the moment—may be thought of as having a degree of group-awarded credits such as to permit idiosyncratic behavior in certain dimensions *before* group sanctions are applied. By definition, affiliation with the group—as perceived by the group—ceases when the individual's credit balance reaches zero.

---

[5] Implicit here is a concern with observable features of the individual. Thus, in the case of conformity, public manifestation is required; another position, taking account of both public and private conformity, has been advanced by Jahoda (Marie Jahoda. "Conformity and Independence: A Psychological Analysis." Unpublished, June, 1956).

It is noteworthy that this concept is applicable to the limited, artificially produced laboratory group as well as to the total society. And, since the individual may have simultaneous membership in many groups, he may be considered to have a distinct credit balance in all groups with which he is in some sense involved; in each case he has achieved some level of status. Affixed to this concept of "credit" is the further consideration that "debits" of varying magnitudes may be charged against the credit balance, depending upon the gravity and frequency of the idiosyncrasy manifested, and the credit level which the individual holds.[6]

Taking our society today as an illustration, one's credit balance very likely will be rapidly exhausted by publicly espousing Communist doctrine. In a different sphere, a fraternity man may experience comparable rejection by his peers for growing a beard, though other factors would come into play, so that for some individuals the consequences—in terms of group sanctions—would be disastrous and for others hardly disturbing. This requires some consideration of factors which determine the awarding of credit.

Among other determinants, the credit balance that a group member achieves depends upon the group, its function, and other properties to be considered below. It is useful for our purposes here to conceive of an "open system," i.e., an autonomous group providing focal activities, as well as free face-to-face interaction yielding expectancies; this would permit the simultaneous observation of an individual's behavior by all group members and the generation of impressions representing credit.

There are three general variables which can be delineated as determinants of these impressions. The first of these is alpha value (V_α), referring to the individual's task competence or performance in regard to focal group activities; the second is beta value (V_β), referring to characteristics of the individual not specific to these activities, e.g., status in a broader group, *bonhomie*, and the like; the third is immediate past idiosyncratic behavior (B), constituting a drain on credits.[7] It is not contended that credit is necessarily related linearly to these variables, nor is their very likely interrelationship ignored. They are doubtless intercorrelated, though of varying degrees of significance in generating or dissipating credits. As a generalization, value (V) tends to increase credit while idiosyncratic behavior (B) acts to decrease credit—though the potential for negative value exists, e.g., in the case of prejudice.

GROUP VARIABLES

From the foregoing it should be apparent that an individual can only be accurate in perceiving expectancies insofar as they are normative, in some modal sense, and are communicated. Two interrelated group variables which have importance in this regard are group attraction (A) and the communicality of any given expectancy (Y).

The former variable may be thought

---

[6] Alterations upward or downward in credit may be conceived as a negative, monotonic function of credit balance. Thus, for the same idiosyncratic behavior or negative weight attached to value, the individual with high status loses less credit than the marginal individual of low status (cf. 20).

[7] Still another variable related to credit balance, probably curvilinearly, would be the duration of the individual's affiliation with the group over time. This has been disregarded, since it is useful to deal with individuals as though they have been in the group for an equal period of time, more particularly from its inception. It is also likely that the degree to which the individual is "visible" may alter the effects produced by his value and idiosyncratic behavior.

of as "cohesiveness," a term more usually applied. But since this term may have at least several operational meanings, it is preferable to specify two kinds of literal attraction, or an aggregation of these: attraction to group members, and attraction to focal group activities. This duality follows the M_a and M_g distinction made earlier in connection with motivation to belong. Although one may deal with a nondiscriminate aggregation of these, it is quite true that the sum of individual attractions in the group may be based predominately in M_a needs or M_g needs, and that differences in the emergent characteristics of groups thus constituted will be evident. Thus, where group attraction derives mainly from M_a one might predict it would be more stable than where its source is mainly M_g, since the latter variable is more temporally based.

Communicality is conceived to be directly related to group attraction in a mutual dependency; it refers to the degree to which an expectancy is literally communicated, i.e., made evident, and bears a relationship to both relevance and communication variables, as they have been introduced in other formulations (cf. 6, 22), though this is by no means a complete statement. Studies of leaders' ability to estimate group attitudes, for example, have yielded highly conflicting results (cf. 21, p. 268). Where attitudes are "relevant," leaders *may or may not* be superior to nonleaders in their respective estimates; where leaders *are* found to be superior to nonleaders in estimating relevant attitudes, this has been ascribed to the heightened social sensitivity of leaders or alternatively to the proposition that leaders are instrumental in the shaping of group attitudes and hence tend to know them better.

Whatever the explanation chosen, there is reason to contend that the variable of relevance may not be the most fruitful one for purposes of study. If one were disposed to test the tenability of the hypothesis that leaders have this greater social sensitivity, it would seem desirable not only to raise the question of whether leaders, and others, tend to use their own personal attitudes as an anchorage for estimating group attitudes, but to ask in addition whether this exists independently of the degree to which given attitudes actually are foci of communication within real groups. Another approach, accordingly, might be to utilize some index of the degree to which a specific attitude—or an expectancy—actually evidences itself in a given group. There is utility, then, in introducing the operationalizable property "communicality."

Though a level of communicality may characterize a group, the particular center of interest here resides in a given expectancy. One may venture in this vein that the communicality of an expectancy will be at a lower relative level than that of other expectancies, if it is less applicable to the group as a whole; common expectancies ought to have higher communicality than the differential expectancies associated with increased status. Since communicality rests on behavior, it may be seen to follow, too, that interpersonal interaction results in higher communicality. Through a related mechanism, interpersonal interaction may contribute to group attraction (cf. 11, 17).

SUMMARY OF VARIABLES

For convenience, the variables described may now be set forth definitionally.

B—Idiosyncratic behavior, i.e., any group member's behavior which may be perceived by the group to deviate from a given group expectancy.

C—Idiosyncrasy credit, i.e., the extent to which a given group member's idiosyncratic behavior (B) is allowable, in terms of gravity and frequency, before group sanctions are applied.

V_α—*Alpha value,* i.e., the weight assigned the current performance of a given individual, which may be perceived by the group as bearing upon its focal activities, e.g., task competence.

V_β—*Beta value,* i.e., the weight assigned the characteristics of a given individual which may be perceived by the group, but are not specific to its current focal activities, e.g., status external to the group.

P_a—The *perceptual ability* of a given individual, in the sense of a capacity to perceive events and relationships in the social field.

P_e—The *perceptual error* of a given individual in perceiving events or relationships in a particular social field, e.g., group expectancies.

M_a—The *motivation* of a given individual to affiliate with a given group, in terms of a gaining or sustaining *social approval.*

M_g—The *motivation* of a given individual to affiliate with a group, in terms of interest in focal *group activity.*

M—Individual motivation to gain or sustain membership, i.e., some composite of a given individual's motivation of both the M_a and M_g variety.

A—In general, *attraction of the group to its members,* i.e., some aggregate of all group members' *M*.

Y—The communicality of a given group expectancy, in terms of the degree to which a given expectancy is evident.

SCHEMATIC REPRESENTATION

In Fig. 1 our symbolic notation has been employed to represent relationships schematically. Since a sequential pattern is of particular importance, a time dimension is involved throughout; thus, subscripts are introduced to indicate the time interval to which reference is made; e.g., t_1 is read as the first time interval; or, P_{e_2} as perceptual error in the second time interval.

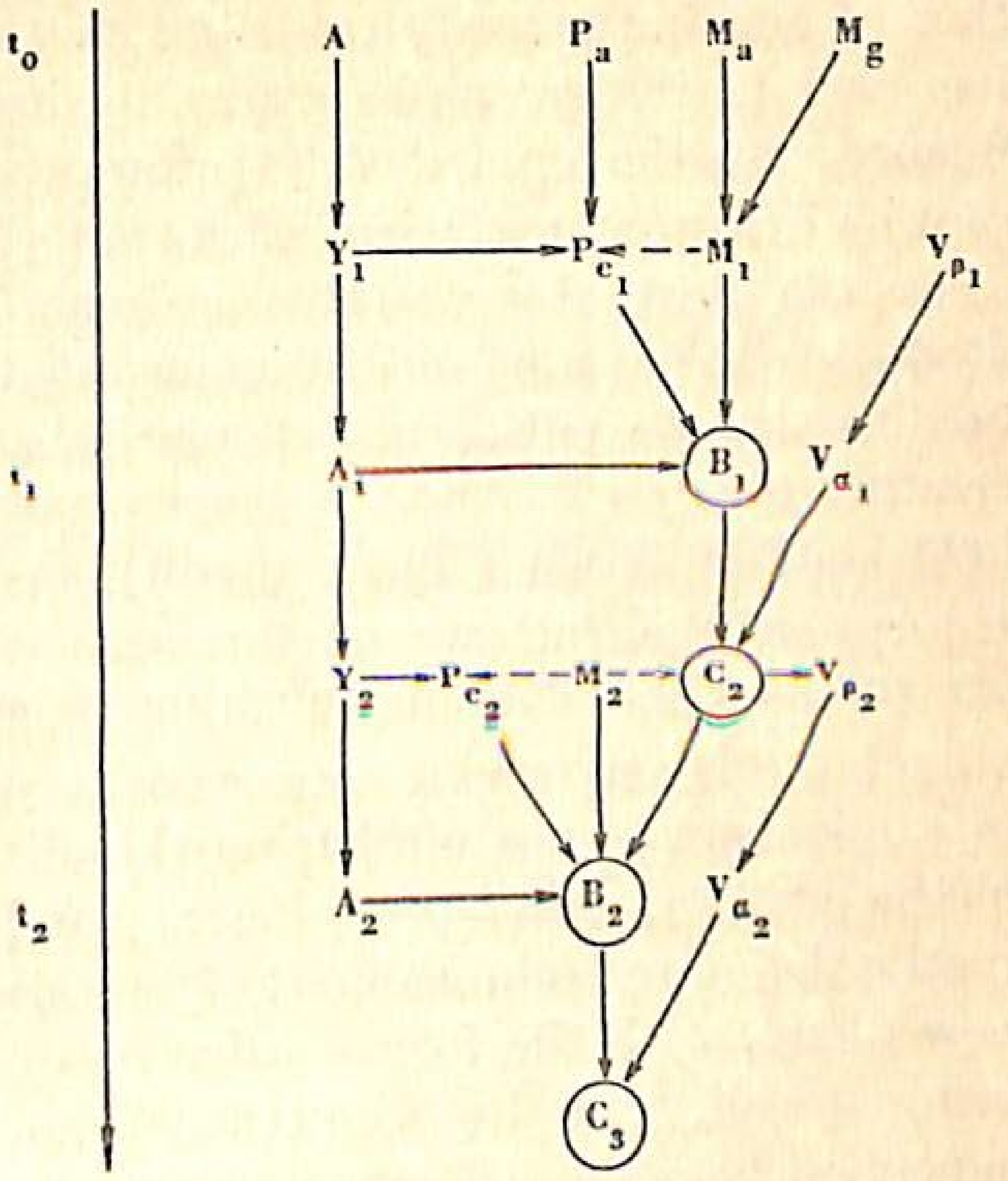


FIG. 1. Schematic representation of mechanisms demonstrating relationships over time.

The system originates at the top with group attraction as a motivational context, and three individual variables, perceptual ability, motivation to gain or sustain social approval, and motivation with reference to the group's activity. At the next level, group attraction has given rise to the communicality of certain expectancies which are then perceived by the individual, thus yielding a perceptual error; and motivation to belong has been aggregated at this level, as well. Beta value is also introduced to signify the group's perception of the individual's characteristics, e.g., pleasant appearance.[8]

Moving down in time, the individual's idiosyncratic behavior during the period just elapsed has been generated by his error in perceiving expectancies together with his motivation—within the constraints imposed by the level of group attraction. The group's perception of the individual's contribution to its focal activities, alpha value, is influenced by the immediately prior perception of his characteristics, beta value.

---

[8] The term "group perception" refers to an abstraction; it is unlikely that all members of a group will perceive a given feature of an individual identically; the intent therefore is only to suggest a modal tendency. However, some very recent research has illuminated just this point and given credence to the general conception of differentially-determined value (cf. 14).

In the next stage, status is generated—in the form of group-awarded credits—by the effects of behavior relative to expectancies and the sequence of beta value to alpha value to credits. At this point the full set of interactions are in play, with credits affecting beta value; the latter serving as a repository of group perceptions of the individual's characteristics; perceptual error and motivation are reintroduced for this new phase, with the former affected by communicality of expectancies. Idiosyncratic behavior is subsequently determined by available credits, as this is checked by motivation, in particular, and perceptual error.

DISCUSSION AND IMPLICATIONS

Early in interaction, conformity to group expectancies serves to maintain or increase status, particularly as it is seen to be combined with manifest contributions to the group; at a later phase, however, the status thus generated permits greater latitude for idiosyncratic behavior. Thus, if an individual conforms to expectancies early in his exposure to the group and if he shows characteristics of competence, he accrues credits. For evident deviations from expectancies, or poor performance, he loses credits. If he exhausts his credit balance completely, pressures are applied to remove him from the group, or, at the very least, he is no longer perceived to be a member. At the other pole, if he continues to amass credits he attains a threshold permitting deviations from common expectancies, but with constraints imposed by newly differentiated expectancies.

The apparent paradox—that leaders both conform to group norms and yet may act to alter them by an exercise of influence—may be explained by reference to this sequential process. In this regard, it should not be supposed that an abundance of credits must lead perforce to influence. While an individual thus endowed has the potential to display more idiosyncratic behavior than others, he might not do so, nor would he of necessity become a leader thereby. Some further points of clarification are in order.

It is easy enough for the individual to continue to do habitually that which is rewarded by relevant others, so long as expectancies remain relatively stable. Consider the state of affairs which holds, however, in the case of the person who has marked status mobility in the group. He cannot simply continue to redisplay behaviors which were appropriate to the group's earlier expectancies, because the expectancies applicable to him are now altered in keeping with his rising status. Other things being equal, this suggests two features appropriate to the attainment of status in an open system: (*a*) accuracy of social perception; and (*b*) modifiability of behavior. Insofar as the incipient status person is attuned to the altering group expectancies and is capable of reacting appropriately to them, his status will very likely move upward. The relationship of these points to research on leadership is noteworthy; for example, the proposition that leaders have a heightened sensitivity to certain properties of the social context is in consonance with the foregoing. Note, too, that whether or not the leader has actually created a feature of the context with which he is then familiar, he may well have evidenced perceptual accuracy in an earlier phase, as he rose to leadership. Accordingly, the finding of Talland (22) that leaders are only better in judging norms where they have had a part in their evolution is not inconsistent with a finding like that of Chowdhry and Newcomb (6) to the effect that leaders have superior social perception; it would appear that the time phase under study is critical in yielding one process or the other.

Previously the point was made that the high-status person could effect changes in the common expectancies of the group because he has latitude for the manifestation of what would be seen to be nonconforming behavior for others. But, in contrast, the expectancies regarding the role itself are less amenable to alteration by the incumbent. Hence, the leader could readily lose credits and find his influence diminished if he were to show idiosyncratic behavior in terms of expectancies associated with his role. Regarding such deviation, we may conceive of one requirement which is quite likely significant, i.e., perception by the members of the group that the leader's motivation to belong be both high and sincere. Should this condition not be fulfilled, status may be threatened. To take another illustration, innovation by the leader may be of high valence to the group. It is conceivable that this could yield the seeming anomaly of a leader who, in the face of this expectancy, adopts a passive and ostensibly safe course, but loses status.

Leadership status, therefore, assuredly demands conformity to the group's expectancies regarding the role, but still leaves the leader with sway in the sphere of common expectancies associated with members at large. The leader may deviate from these, or bring about their reconstruction, if his prior activities have generated an appropriately high level of credit.

The motivational aspects of this process require consideration in terms of the individual set against the background of the group's activity. The restraining effect of M on the expenditure of C has been accounted for in the foregoing. What is more to the point, however, is the fact that a person with M_g will more likely achieve status in a largely M_g group than will a person with M_a; the status achieved, therefore, is a part function of the congruence of the individual's motivation with the generalized character of the motivation extant in the group. Since motivation is related, as well, to performance and other characteristics represented by alpha value and beta value, it is to be regarded as a key element.

If the group has a primary focal activity, then presumably M_g becomes more highly valued, particularly in combination with alpha value. At the other extreme, i.e., where the group activities are quite diffuse, M_a becomes important in combination with beta value. We may conceptualize groups of the M_a variety as being essentially "socially minded." With a minimum group-centered function requiring broad participation, the person having beta value is more likely to achieve status through being well liked; in the M_g group, to the contrary, alpha value becomes critical and task competence has greater weight in determining status. The relationship of this to the current situational view of leadership is evident.[9]

Whether group members do distinguish between value of one kind or another can be inferred from the literature of sociometry. Evidence will be found there indicating that individuals can give scaled evaluations of their peers with quite adequate discriminations between those they like, those they consider competent, and so forth. Though these sociometrically based status continua are likely to be related, as we have previously noted, they are by no means in universal, one-to-one correspondence (cf. 10). Viewing status in the aggregative, credit-amassing sense still allows for the integrity of the roots which feed it. Recent experimental evidence on the basis for shifts in the group's perception of a member also accords with this conception (cf. 14).

---

[9] For example, the finding of Jennings (13) that leadership and popularity are highly related in her groups can be considered in this framework.

Certain of the assumptions made here—e.g., that an individual will have a level of credit reposing unitarily in others' perceptions of him, and that he may know and make use of the credits at his disposal—are only approximations of reality. Their literal tenability is not crucial to the mechanisms postulated, however. One could argue that the individual operates *as if* these assumptions were in fact true: the "they" commonly invoked to denote the upholders of some social pattern are never quite as homogeneous as the term suggests; but, to the individual, the use of "they" to represent a supposed uniformity is a necessary convenience as a basis for behavior. Furthermore, in accordance with this position, the individual apparently does react differentially to what he believes to be the view of him held by the "they," as Dittes and Kelley (7) have demonstrated by manipulating the level of group "acceptance" which an individual is permitted to sense. In general, then, it appears that the individual seeks to know where he stands and does the best he can with the information available to him. These conceptions therefore do no violence to the reality with which the individual deals, but rather describe this reality in terms congruent with his concern.

SUMMARY

Beginning with the consideration that social behavior depends upon attributes of the individual, conditions of the situation, and inputs to a dynamic system arising from their interaction, a theoretical conception relating conformity and status is presented.

The major mediating construct introduced is "idiosyncrasy credit," taken to be an index of status, in the operational sense of permitting deviations from common "expectancies" of the group.

Credits are postulated to increase or decrease as a function of the group's perception of the individual's task performance and generalized characteristics, and of his "idiosyncratic behavior," i.e., deviations from its expectancies. Though increases in credit are seen to permit greater latitude for idiosyncratic behavior, motivational and perceptual states of the individual, and group-level phenomena, are also considered.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BATES, F. L. Position, role and status: A reformulation of concepts. *Soc. Forces,* 1956, **34**, 313–321.
3. BENDER, I. E., & HASTDORF, A. H. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.,* 1953, **48**, 503–506.
4. BOVARD, E. W. Conformity to social norms and attraction to the group. *Science,* 1953, **118**, 598–599.
5. BRUNER, J. S. On perceptual readiness. *Psychol. Rev.,* 1957, **64**, 123–152.
6. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.,* 1952, **47**, 51–57.
7. DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.,* 1956, **53**, 100–107.
8. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.,* 1955, **51**, 629–636.
9. FESTINGER, L. Informal social communication. *Psychol. Rev.,* 1950, **57**, 271–282.
10. HOLLANDER, E. P., & WEBB, W. B. Leadership, followership, and friendship: An analysis of peer nominations. *J. abnorm. soc. Psychol.,* 1955, **50**, 163–167.

11. Homans, G. C. *The human group.* New York: Harcourt, Brace, 1950.
12. Jackson, J. M., & Saltzstein, H. D. *Group membership and conformity processes.* Ann Arbor, Mich.: Research Center for Group Dynamics, 1956.
13. Jennings, Helen H. *Leadership and isolation.* (Rev. ed.) New York: Longmans, Green, 1950.
14. Jones, E. E., & deCharms, R. Changes in social perception as a function of the personal relevance of behavior. *Sociometry,* 1957, **20,** 75–85.
15. Krech, D. Psychological theory and social psychology. In H. Helson (Ed.), *Theoretical foundations of psychology.* New York: D. Van Nostrand, 1951.
16. Newcomb, T. M. *Social psychology.* New York: Dryden, 1950.
17. Newcomb, T. M. The prediction of interpersonal attraction. *Amer. Psychologist,* 1956, **11,** 575–586.
18. Postman, L., Bruner, J. S., & McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.,* 1948, **43,** 142–154.
19. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.,* 1948, **43,** 259–278.
20. Schachter, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.,* 1951, **46,** 190–207.
21. Steiner, I. D. Interpersonal behavior as influenced by accuracy of social perception. *Psychol. Rev.,* 1955, **62,** 268–274.
22. Talland, G. A. The assessment of group opinion by leaders, and their influence on its formation. *J. abnorm. soc. Psychol.,* 1954, **49,** 431–434.
23. Thibaut, J. W., & Strickland, L. H. Psychological set and social conformity. *J. Pers.,* 1956, **25,** 115–129.

(Received January 30, 1957)

ERRATUM

In the article "Uncertainty and Conflict: A Point of Contact Between Information-Theory and Behavior-Theory Concepts" by D. E. Berlyne in the November 1957 issue of this Journal (*Psychol. Rev.,* 1957, 64, 329–339), there are two errors. On p. 332, in Property 5, "$\equiv E_1$" should be substituted for "> 0," and on p. 333, Expression 2 should read "$\Sigma_i E_i(\log(\Sigma_i E_i) - \log E_i)$."

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by JAMES DREVER

Vol. 49 Part 1 February, 1958 20s. net

H. P. HILDEBRAND. A factorial study of introversion-extraversion.

P. L. BROADHURST. Determinants of emotionality in the rat. II. Antecedent factors.

ROBERT T. GREEN. Surprise, isolation, and structural change as factors affecting recall of a temporal series.

L. S. PENROSE and R. PENROSE. Impossible objects: a special type of visual illusion.

MARY WICKHAM. The effects of the menstrual cycle on test performance.

N. O'CONNOR and G. S. CLARIDGE. A 'Crespi effect' in male imbeciles.

D. WALTON and T. L. BEGG. The effects of incentives on the performance of defective imbeciles.

P. L. BROADHURST. A 'Crespi effect' in the analysis of emotionality as a drive in rats.

M. D. ALLAN. A pattern recognition method of learning Morse code.

P. N. R. SUTTON. The development of independent muscular control of separate eyelids in two racial groups: European and Polynesian.

J. A. DEUTSCH. Shape recognition: a reply to Dodwell.

PUBLICATIONS RECENTLY RECEIVED.

OTHER PUBLICATIONS RECEIVED.

The subscription price per volume, payable in advance, is 60s. (or $10) net (post free).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS
Bentley House, Euston Road, London, N. W. 1

BUREAU OF EDNL. & PSY. RESEARCH LIBRARY

VOL. 65, NO. 3 MAY, 1958

THE PSYCHOLOGICAL REVIEW

ACTION DECREMENT AND ITS RELATION TO LEARNING[1]

EDWARD L. WALKER

University of Michigan

It is the purpose of this paper to set forth some argument and some evidence that what is herein called *action decrement* and what has otherwise been labeled *reactive inhibition* (I_R) or *stimulus satiation* (${}_SI$) is, at least in some instances, *positively* related to learning. This is in contrast to the usually held position that it is *negatively* related or antithetical to learning. Furthermore, it will be argued that when one tracks the course of the action decrement experimentally, one also, in fact, tracks the course of the perseverative consolidation which is frequently assumed to follow a period of training and is assumed to be necessary for subsequent performance.

The use of the term *action decrement* rather than any of the other available terms stems from the following beliefs. *Reactive inhibition* and *stimulus satiation* are no longer considered suitable labels because of the growing belief that any psychological action or event is followed by a temporary restriction upon its rearousal. It no longer seems certain that there is an exclusive relationship between action decrement and overt responses or to stimulus events, although it is admittedly difficult to deal with psychological events which are not stimulus-aroused. In short, the position taken is that the action decrement is a central event and the relevant symbol would thus be $(\underset{c}{I})$.

The contrasting points of view with respect to the role of action decrements can be elaborated on both a common-sense and on a theoretical level. On a common-sense level it is reasonable to view fatigue and fatigue-like states as biologically useful in preventing the organism from exceeding its resources and thus suffering damage. Such protection could be regarded as a sufficient function. Also on the common-sense level, we all know, or think we know, that a tired organism learns very inefficiently. Thus common sense tells us that Hull's (5, 6, 7) projection of these antithetical and detrimental effects to the theoretical level is eminently reasonable.

However, common sense sometimes provides contradictory propositions. It is possible to say that learning is not likely to occur without practice. Much practice is associated with much fatigue. The more one practices, the more one learns. Therefore, the more fatigue, the more learning. On a formal logical basis the conclusion is not certain, for there is a logical fallacy in the sequence of premises. Therefore, empirical demonstrations are required to establish the

---

[1] Most of the work reported in this paper has been made possible by a grant from the National Science Foundation.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated.............................

truth of the conclusion. Several will be cited subsequently in this paper. Furthermore, the protective character of the fatigue-like state is retained, although the object of protection is a perseverative process rather than the whole organism.

The concept of perseverative consolidation in learning is an intriguing one. The concept in its most general form is quite simply stated. When an organism practices an action, there continues for a time a process which produces consolidation of the effects of the practice. If this consolidation process is interrupted or interfered with, by such events as a traumatic blow on the head or by practice on other similar material, as in the retroactive-inhibition paradigm, subsequent performance of the act will be deficient as a result of the interference. It is hypothesized that the action decrement is a direct result of the process which produces the perseverative consolidation. The biological utility of fatigue-like states, in this instance, lies in their capacity to protect the consolidation process from interference. A limited amount of practice, in any given instance, is efficacious. Beyond that limit fatigue is accompanied by a modification of the character of the process into some undesirable or inefficient form. Such modifications would constitute interference with the desired response. The fatigue-like state permits as much efficacious practice as the organism can tolerate without essentially modifying the character of the action system and its associated behavior.

Sharpless and Jasper (12) point out that there is a variety of decremental effects which occur as a function of continued or repeated stimulation. They make several distinctions. Among them there is a clear difference between "habituation" and "effector fatigue." For them,

> Habituation, like learning, is clearly referable to some form of plasticity in the central nervous system. . . . It does not depend on effector fatigue, since a response which has become habituated to a specific stimulus may still be elicited by an appropriate novel stimulus, whereas a fatigued effector will fail to respond, not only to the repeated stimulus, but to all other stimuli of the same relative intensity.

"Action decrement" is used in this paper to mean something very similar to "habituation" in Sharpless and Jasper. It is something like effector fatigue, which will accumulate in sufficient quantities to eventually modify the character of the response, usually in some undesirable form. Action decrement serves to delay or inhibit reactivation and thus to protect the consolidation process from the accumulation of effector fatigue to the point of serious modification of the character of the action system. Effector fatigue may be only one of several mechanisms of interference and may be chiefly associated with overt responses. Effector fatigue is not thought to be a significant factor in any of the situations considered in this paper.

The theory can be restated in a more concise, although not as yet in a very elegant form. *Any psychological action is followed by an action decrement—a lowered capacity for rearousal of the same event. The action decrement is a direct manifestation of the process of perseverative consolidation which is necessary for retention and subsequent performance. The action decrement persists for a limited time and then dissipates. Under many circumstances the dissipation of the action decrement is followed by an action increment which is learning or habit strength.*

The argument at this time is limited in the following sense. An attempt will be made to present positive, empirical evidence for the theory, but no attempt will be made to account for every phenomenon in every situation in which concepts such as reactive inhibition and conditioned inhibition have been effectively and powerfully applied as negative factors, or factors antithetical to positive learning.

THE MODEL

What follows is a verbal model which has the potentiality of becoming, under intelligent modification, a rigorously deductive model, a neurophysiological model, a mathematical model, or all three at once. It is not clearly any of these in its present form.

Perhaps the best introduction is to state the problems inherent in what has already been said. It is fairly common practice to think of perseverative consolidation in terms of reverberatory circuits or systems. Reverberatory systems have two basic deficiencies, from the present point of view. There are grave suspicions that reverberatory activity with the initial stimulus as the sole impetus could not last as long as the consolidation process seems to last. Furthermore, one would expect such perseveration to present a condition in which it is *easier* to rearouse the action, rather than more difficult. That is, if some fraction of an action system continues in a state of reverberation, the threshold for that action should be lowered. It has been hypothesized here that the consolidation process is indexed by a decrement, a raised threshold.

Neurophysiological literature contains many empirical examples of the aftereffect of stimulation in systems more complex than a single fiber. One example which seems to offer promise may be seen in Fig. 1.

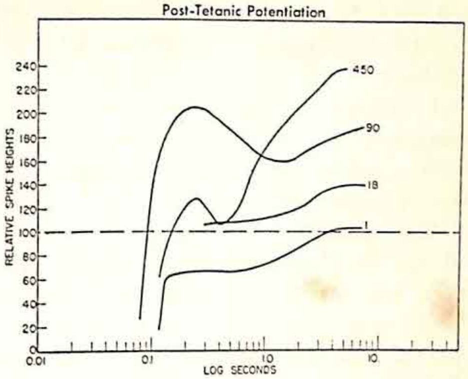


FIG. 1. Changes in sensitivity as a function of time after stimulation in a monosynaptic reflex system. (Redrawn on a log scale from Eccles and Rall, 1, 2.) The numbers 1, 18, 90, and 450 refer to the number of conditioning pulses delivered to nerve to the Flexor Longus Digitorum at a rate of 300 per second. Ordinate values represent height of test spike relative to first spike from conditioning stimulus.

This figure is redrawn from *The Neurophysiological Basis of Mind* (1) by J. C. Eccles. The data in Eccles are taken from a study reported by Eccles and Rall in 1951 (2). A monosynaptic reflex system is prepared in such a manner that the nerve trunk afferent to the reflex can be stimulated, and recordings of potentials can be taken on the efferent side of the synapse. The spike height of the first impulse is taken as the baseline (100%). The effects of any given number of conditioning pulses can then be traced by sending in a test pulse after a controlled delay. The height of the test spike is then expressed as a proportion of the first spike produced by the conditioning pulse. In the curve labeled "1" it can be seen that the second pulse was less than the first, if the test pulse was delivered any time during the first three seconds after the conditioning pulse. Thus there was an output decrement traceable for three seconds or more. If the conditioning stimulus involved 18 pulses, in this case delivered at a rate of about 300 per second, no decrement is visible in the graph; but there is a period of postsynaptic potentiation lasting for at least eight seconds. In the curves for 90 and 450 conditioning pulses, very brief decrements are plotted, followed by increments of 150% or more.

There are many reasons why it is dangerous to make much use of these data: these are motor neurones at the spinal cord level, the conditioning stimuli are not physiological stimuli, they amount to punishment of the fiber in a physoiological sense, etc. These data are offered only to support the reasonableness of the following propositions.

1. It is possible that the activation of neural tissue results in a decrement in output and sensitivity followed by an increment.

2. It is possible that the more intense the activity, the quicker the recovery, the greater the rate of recovery, and the greater the increment or potentiation.

These descriptive statements—of successive states of the neural tissue involved in a particular action and of the results of variation in intensity—then form the basis of a model.

When an animal alternates choices between two trials in a **T** maze, the phenomenon can be attributed to a decrement in the action system involved uniquely in the first choice. The raised threshold for this action forces the choice of the other alternative on the second trial. This situation can be used to illustrate and develop the model further. An animal is placed in the maze, turns left at the choice point, and is then removed. Only a small portion of the neural tissue available is assumed to have been involved in the action of choice, and it seems likely that many of the units so involved are fired several times in the process. It is possible to conceive that these units go into a decremental phase together, and in a brief period, possibly a second or two, recover more or less together, and then rise into a potentiated or supersensitive phase together. This picture would fall somewhere between 1 and 18 on the Eccles and Rall graph.

At this point let us invoke the Magoun Ascending Reticular System and its nonspecific cortical input. It is possible to conceive of either a rain of nonspecific impulses or of rhythmic waves of arousal activity at EEG frequencies as bringing about a refiring or recycling of the system after it rises into its supersensitive phase. It is obvious that such a mechanism could involve only a fraction of the original action system, for the animal does not repeatedly turn left at intervals from a fraction of a second to one or two seconds.

Assuming that only a portion, but a necessary portion of the system, goes through the process, the picture might be something like that seen in Fig. 2A, in

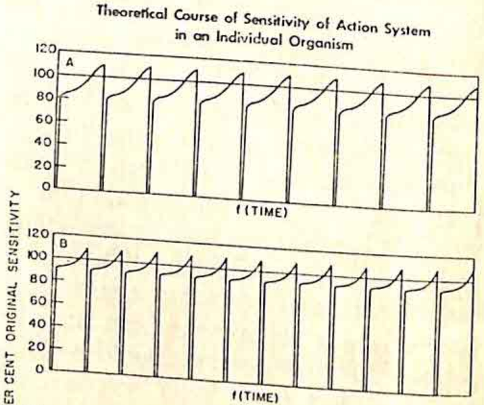


FIG. 2A. Theoretical recycling pattern for an action system driven by nonspecific input originating from a low drive state.

FIG. 2B. Theoretical recycling pattern under a higher drive state.

FIG. 2C. Theoretical recycling pattern under a low drive state with reward introduced after the third cycle. See text for explanation.

which the choice response is assumed to have occurred at the left. The fraction of the system goes into a decremental phase, and recovers into a potentiated phase. When it reaches an appropriate level, nonspecific input originating in some part of the reticular system can become effective in rearousing a fraction of the original action system, which then goes into a second decremental phase. This iterative process might well recur many times.

A further step involves the effects of different levels of motivation. Suppose one were to assume, as does Hebb (4), that differences in motivational level correspond in some way to differences in nonspecific output from the reticular system. According to the descriptive statement relative to the effects of intensity drawn from Eccles and Rall, one would expect an increase in the level of nonspecific input from an increase in motivational level and thus an increase in the intensity of the original action or any subsequent cycle of it. This would produce a quicker recovery in absolute time, a steeper curve of recovery, and a greater potential increment. Figure 2B illustrates the effect of a higher level of motivation on the shape of the cycle and the rate of iteration.

A final step involves an assignment of a role for reward within the model. Consideration leads to the conjecture that any reward condition might produce an increase in the level of nonspecific input, very much as if the level of motivation had been raised to a higher value. This condition is pictured, theoretically, in Fig. 2C, where reward is assumed to have been introduced after the third cycle. This effect of reward would fit the results of brain stimulation studies as reported by Olds and Milner (11), and as summarized by Olds (10). An implanted electrode with its tip in various parts of the brain can produce approximately the same effect on behavior as does water reward to thirsty animals. Both are conceptualized here as producing increases in the level on nonspecific input.

There are at least two factors which might cause this recycling process to cease. The integrity of the action system is dependent upon the condition that all units are originally fired at approximately the same time. At each recycling, a certain number of units could be expected to be lost because they did not recover in time to participate in the next recycling. Thus the size of the action system would grow progressively smaller through a process of fractionation. Furthermore, such a system has to some degree an afferent and an efferent end. One might conceive of the recycling process as being progressively restricted to the more efferent end. Both factors could be expected to produce a progressive decrease in the correlation between the excitability cycles of the units of the action system. The system would be conceived to have lost its dynamic integrity, and no further consolidation would occur from the moment when the correlation between the units reaches zero. When a unit drops out of the system, it can be expected to recover rather quickly to a semipermanent state of slightly lowered threshold, and when all perseverative activity has stopped, the combined threshold changes will constitute an increment in habit strength.

Certain characteristics of the model should probably be noted at this point. It describes a hypothetical pattern of behavior for a population of neural units which involve time units of the order of seconds. It may eventually be possible to derive this model from one which describes the behavior of individual units in time units of the order of milliseconds. Second, it will be noted that the burden of perseveration

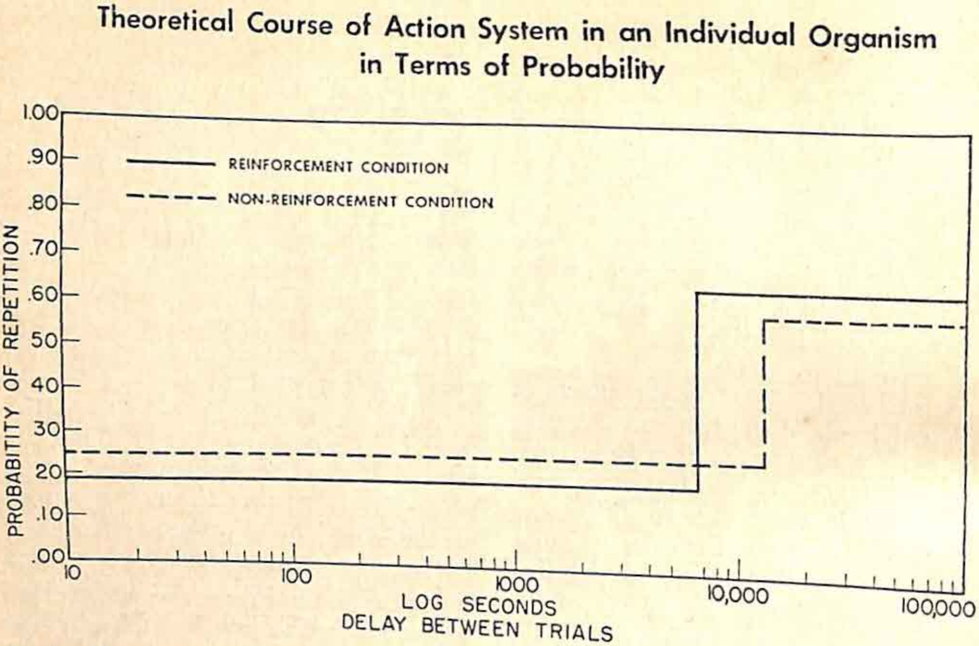


FIG. 3. Calculated probabilities of repetition for a nonreinforcement condition (from Fig. 2A), and from a reinforcement condition (from the latter part of Fig. 2C).

is in large part carried by the states of the units involved, where the state is a function of the original condition and the effect of the action. Thus, while perseverative recycling might be expected to be wiped out by electroconvulsive shock, it might well survive cooling in a hibernating animal. Cooling might slow or even stop the recycling process, but subsequent warming might bring it back into a perseverative condition. Finally, the model at this point is largely deterministic. At the next most molar level it becomes a probabilistic model.

Probabilities become factors in the model in several ways. As an action system rises into a potentiated phase, the precise point of recycling is subject to more or less random variation. Figure 2 might well have been drawn with different degrees of potentiation before recycling. Had this been done, there would have been an irregularity in the heights reached by the individual cycles, and there would have been a resulting variation in the time consumed by each cycle. Much more important, however, is the problem of the state of the system at the moment when the organism makes a second choice. If the system is in decrement, then the organism should choose the previously unchosen alternative. If it is in an incremental or potentiated phase, the choice should be to repeat the original choice. Since the time intervals involved in a single cycle in Fig. 2 should be in the general range of a second, and since experimental situations employed so far do not permit time control with any precision at this time range, the state of the system on the occasion of a second choice should be essentially a random sample of the available conditions. The prediction of what an animal will do on the occasion of a second choice can only be expressed in terms of probabilities.

The probability of repetition should be quite close to the ratio of time in a decremental state to time in a potentiated state. In Fig. 2A this ratio is 150:50, yielding a probability (p) of repetition of .25. In Fig. 2B and in the latter part of Fig. 2C, the ratio is 163:37, yielding a p value of .185. Thus the initial p values plotted in Fig. 3 are derived directly from Fig. 2.

In Fig. 3 the assumption is made that as long as any part of the action system is in an active state, as portrayed in Fig. 2, the probability of repetition will remain at a fixed value. When the active phase comes to an end, this p value will rise to a level slightly higher than it was before the activation of the system occurred originally. Obviously, repeated activation will lead eventually to a p value of nearly 1.00.

It is interesting to note that Fig. 3 conforms to the hypothesis previously stated by Walker (14) that the *greater* the decrement, the quicker the recovery and the greater the eventual increment (learning). In Fig. 1 the quick recovery and the greater increment are associated with the *smaller* apparent decrement. Thus one of the three correlated propositions is inverted in the process of transition from the deterministic level to the probabilistic one.

It is tentatively assumed that the amount of the decrement, the time required for recovery from the active phase, and the height of the increment will all be proportionally related. Whatever influences one will have a proportional effect on the other two. This should not only be true of experimentally manipulated variables such as level of motivation and amount of reward, but should describe individual differences between organisms as well. That is, if, under the same experimental conditions, two organisms show different probabilities of repetition during the decremental phase, the one with the lower probability of repeating should

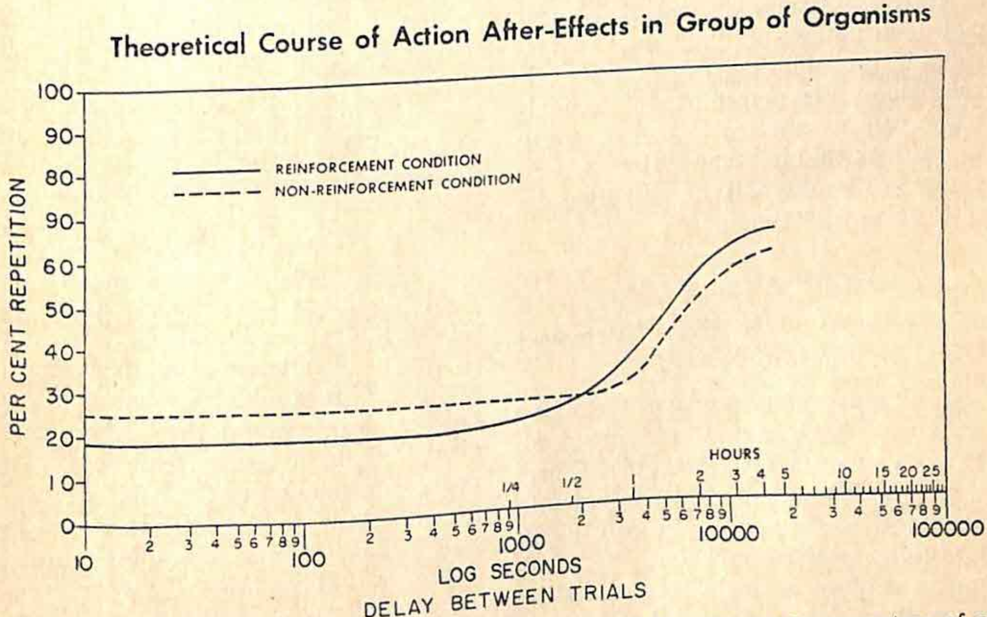


Fig. 4. Theoretical course of action aftereffects, expressed in terms of the percentage of a group of organisms which can be expected to repeat (or alternate) their first choice when given a second choice after a controlled delay.

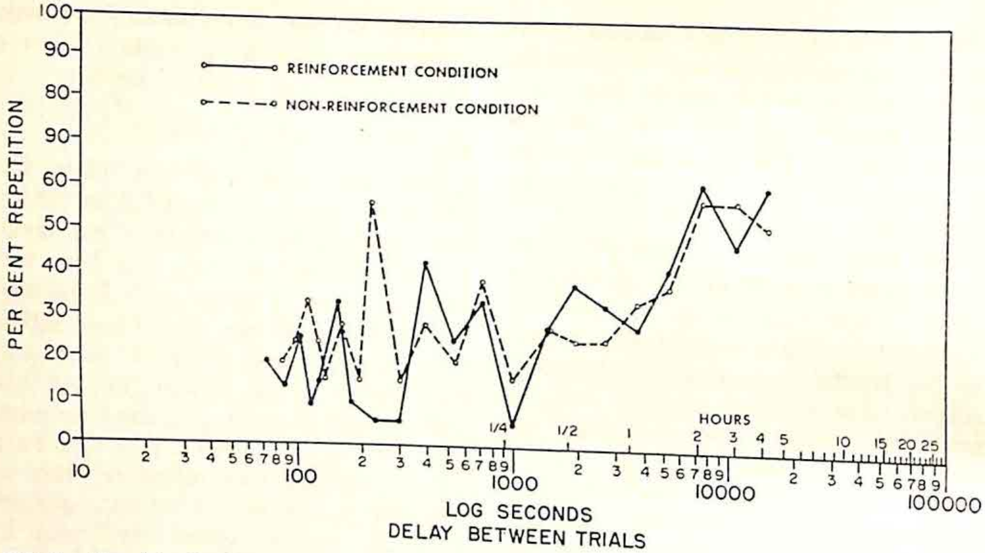


FIG. 5. Empirically determined course of action aftereffects, taken from Walker (13). Figure 4 is an approximate smooth fit to these data.

show quicker recovery and a greater increment of habit strength from a single trial. Since there is no immediately apparent way to assess the probability of repetition in a single organism, it is necessary to assume an essentially normal distribution of individual differences in probabilities; this permits calculation of the percentage of animals which can be expected to repeat at any given moment in time after a first choice of two alternatives. Figure 4 depicts this expectation.

EMPIRICAL EVIDENCE

The Duration of the Action Decrement and the Effect of Reward

The duration and course of the action decrement have been traced in an alteration study in a T maze (14). There were two groups of thirsty animals. One received water in both goal boxes; one group received no water. Each animal was given a single pair of choices on a given day, with the time between choices carefully controlled. Twenty-one different time intervals, ranging from about 80 seconds between trials to something over four hours, were investigated; there were 21 animals per point. The model, as specified, predicted two things.

First, it had previously been thought that action decrement, at least under its other names, would dissipate as an exponential function of time. The model suggested that the decrement would be sustained for a period without appreciable loss as a function of the recycling process, and then recovery would occur.

Second, when the reward and nonreward conditions were compared, it was expected that there would be a greater tendency to alternate in the reward group. This would be followed by a quicker recovery and then, after sufficient time, by a greater tendency to repeat, as a function of the habit strength acquired in a single trial. It should be noted that there is something a little nonsensical about this. A greater tendency to alternate in the rewarded animals would mean that if two thirsty animals turn left on the first trial, and if one is rewarded with water and the other is not, the *rewarded* animal should have a greater tendency to turn the other way on the subsequent trial. But on the basis of acquired habit strength, alone, it should have a greater tendency to repeat its first choice at all times. The results may be seen in Fig. 5.

The expectation of a sustained decrement seems to be borne out by these results. There is a greater tendency to alternate (or lesser tendency to repeat) in the rewarded animals in the early stages. The quicker recovery for the rewarded group is suggested, but no more, and the habit strength from a single reward could not be demonstrated with the time and resources available. The theoretical curves in Fig. 4 are actually a smooth fit to the data in Fig. 5. The general form of the two curves and the relation between them were established prior to the experiment, but there was no way to suggest appropriate repetition percentages or time values.

These results seem to provide partial support for the theory and are certainly somewhat difficult to explain from a more traditional point of view.

Motivation and Alteration

The effort to deal with motivation has not been quite as successful as the effort to deal with reward. One previously unpublished study has been carried out using a shock grid throughout a T maze, with the goal boxes shock-free. A high level of shock (40 volts) was applied to the animal (his feet had been dipped in saline solution) throughout two trials. It was hoped to show a greater tendency to alternate than in the thirst and reward study, quicker recovery, and a greater increment in habit strength. There are only four points plotted in Fig. 6. The decrement is greater than in the previous study, but the recovery rate seems faster, since both of the middle points are out of the population of points in the thirst-reward study. The final point, at 24 hours, does not constitute a sig-

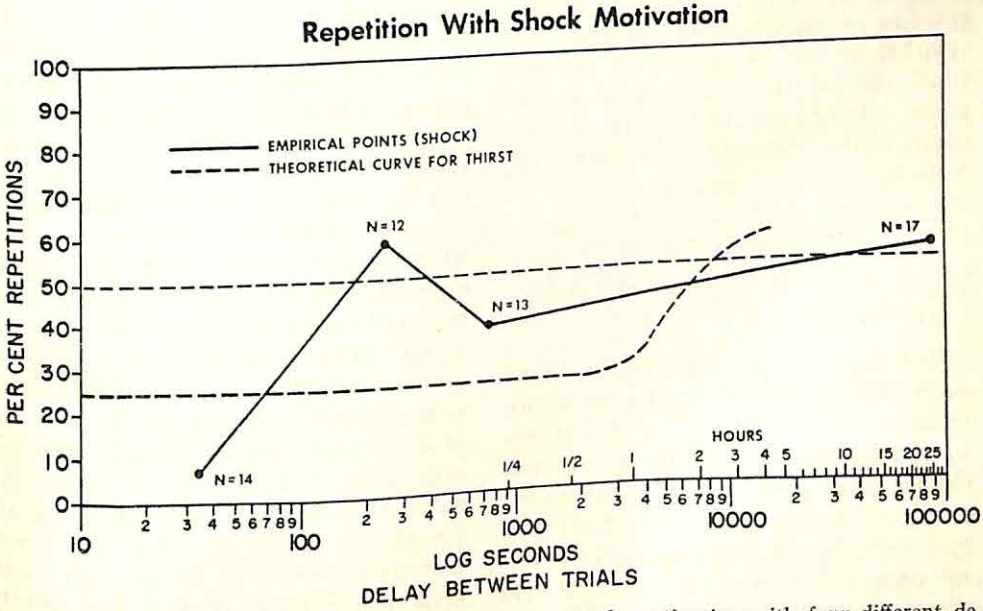


Fig. 6. The tendency to repeat choices under strong shock motivation with four different delays between trials, compared with the theoretical curve for strong thirst (see Fig. 4).

nificant demonstration of habit strength from a single trial.

Amount of Decrement and Ease of Learning

A very different kind of study emerges from a restatement of the theory in the following way. *Any factor capable of producing a big decrement should be capable of producing a big increment or rapid learning.*

The background for this experiment by Walker and Paradise (17) begins with a study by Glanzer (3), and a nearly identical one by Montgomery (9). Both reasoned that when an animal alternated in a T maze, one could not tell whether it was now turning right because it had previously turned left, or whether it now went to the white side because it had tried the black side the last time. Both built + shaped mazes that could be used as T mazes from either top or bottom by closing off one starting alley. If an animal is given a trial from one end and turns left to the black, and if his second trial is from the other starting alley, he can now alternate responses and turn right to the black, or repeat his left turn and alternate stimuli to go to the white, but he cannot alternate with respect to both stimulus and response. Both Glanser and Montgomery found alternation with respect to stimuli, and virtually none with respect to response.

Subsequently, Walker, Dember, Earl, and Karoly (15) went a step further, placing the + maze on wheels; with a somewhat more complex design they were able to offer animals a choice of alternation with respect to all three—stimulus, place, and response, or any one against the other two. It was found that stimulus was most important, and, by implication, was producing the greatest action decrement; place was next most important; and response was last.

Following that study, it was reasoned that the response could be made a significant factor in alternation if the feedback from the two responses was qualitatively very different. To accomplish this, a slightly modified + maze, called the side-winder or the contortion maze, was built. The animal was required to twist in three dimensions to enter the goal alley, and the responses to the two sides required two very different movement patterns. The previous design was now repeated by Walker, Dember, Earl, Fawl, and Karoly (16), and it was found that the response was now a considerable factor in alternation.

Since three factors could be combined or separated in the original (horizontal maze, H), and since there were four estimates of the differential capacity to produce alternation in the side-winder (vertical maze, V), there were available eight conditions which differed in their capacities to produce alternation. By inference, they differed in the relative depths of action decrement they produced. By hypothesis, they should differ in the same order in their efficacy as cues in learning. Therefore, eight learning conditions, four in each maze, were set up. In one condition in each maze, the reward remained with the same stimulus, the same place, and the same response, all at the same time (SPR-H and SPR-V). In another, the reward followed the stimulus, but was random with respect to response and place (S-H and S-V). In the third, the reward remained in a place, but was random with respect to the stimulus and response (P-H and P-V). In the fourth, the reward followed the response, but was random with respect to the stimulus and place (R-H and R-V). The results, with eight animals in each group, are shown in Fig. 7. The top figure shows the percentage of repetition (the reverse of alternation) found to each of the eight factors in the previous studies. Thus the letters SPR-V indicate that,

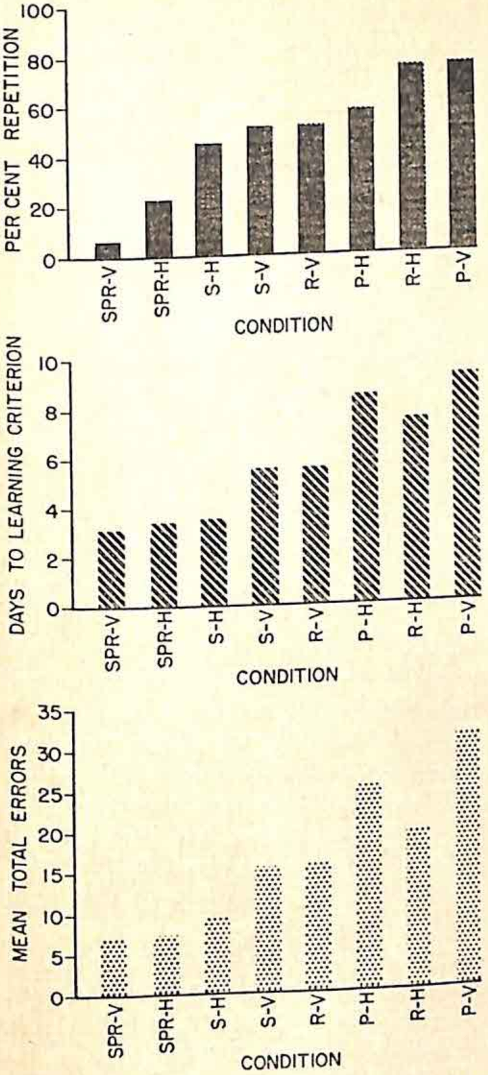


FIG. 7. Empirical test of proposition that the amount of decrement is positively related to rate of learning. Data taken from Walker and Paradise (16). See text for explanation of conditions.

when the animal was permitted to alternate with respect to all three in the vertical maze, there was very little tendency to repeat a choice but a great tendency to alternate, and thus a very large action decrement. On the other hand, the letters P-V mean that in the vertical maze, when the animal was offered an opportunity to alternate stimulus and response by repeating place, there was a strong tendency to repeat place, and, by inference, little action decrement.

The middle figure shows the number of days required to reach a criterion of learning in each of eight conditions. There are eight trials per day under thirst with water reward, and the criterion is 15 correct choices in 16 trials. The lower figure indicates the mean number of errors in each group. The fit is reasonably close and, to the extent that it is close, supports the theory.

Action Decrement, Stimulus Trace, and Perseverative Consolidation

It was asserted earlier that the process being tracked when one measures action decrement is perseverative consolidation. It also seems essentially true that a stimulus trace and a memory trace involve essentially the same mechanism. It was for this reason that the basic hypothesis of this paper was cast in terms of "any psychological action." This process consists of a progressive loss of members of the original population of neural units involved in the action.

According to the present hypothesis, the occurrence of the event in question, whether a presentation of a conditioned stimulus or a learning trial, serves to correlate the sensitivity cycles of a population of neural units. Driven by nonspecific input, a part of the system thus created goes through a neurally active, iterative, recycling process. At each cycle some of the original population of neural units is lost. The rate of loss will determine the duration of the active phase and the form of any measure of the effectiveness of the trace at any given point in time.

The form of such a trace is certainly a "simple decay function," as suggested by Hull (5) in 1943, and in certain instances is well fitted by a power func-

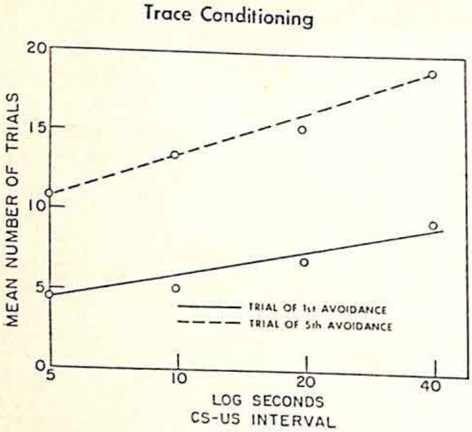


FIG. 8. Two measures of rate of learning of a trace-conditioned avoidance response with four different CS-US intervals. Data taken from Kamin (8), and drawn on a log abscissa.

tion, as indicated by Hull (6) in 1950. Kamin (8) has recently reported data on trace conditioning in dogs. He used CS-US intervals of 5, 10, 20, and 40 seconds, and reported a variety of measures of learning and extinction as a function of the duration of the trace. Certain of his measures of rate of learning seem to approximate a straight line when plotted on a log abscissa, as may be seen in Fig. 8. These particular measures were selected from several reported by Kamin, since it seemed likely that they were least affected by the fact that two animals in the 40-second group failed to learn to criterion.

What appear to be the least ambiguous results on the rate of decay of a memory trace or learning trace are those reported by Thompson and Dean (13). Following a learning trial, each recycling of the action system is conceived of as contribution to the habit strength acquired on that trial in proportion to the number of the original neural units remaining in the system in that cycle. If one should interrupt the recycling process at any given time after the learning trial, the amount of habit strength laid down by that trial should be describable as a linear function of log time between the trial and the interruption.

Unfortunately, it is extremely difficult to demonstrate the quantity of habit strength acquired on a single trial, and therefore even more difficult to demonstrate that a portion of that amount was lost through a given interruption. However, training can be massed so that one trial follows another in rapid succession and a large number of trials forcibly reactivate the action system without appreciable decay between trials. Under these conditions, a sufficient quantity of potential habit strength can be built up so that one would expect demonstrable differences in perseverative consolidation as a function of the interval between the end of training and the interruption. If the memory trace behaves as does the stimulus trace in the Kamin study (Fig. 8), then the amount of habit strength should be a linear increasing function of log time. Since the action decrement seemed to be virtually at an end after four hours in the Walker study (14), the theoretical curve in Fig. 9 is plotted to that point (see Figs. 4 and 5).

The study by Thompson and Dean (13) has the following character. They used a very simple maze with a starting box and a small V shaped choice chamber with two windows at the end, separated by a small board projecting out three inches from the window wall. They prehandled their animals for several days, and then, in preparing them for the critical learning period, taught them a simple grayness-discrimination problem for three days. They were motivated by shock applied if they stayed in the starting box more than 5 seconds or if they failed to choose a window after more than 30 seconds. There is no statement of how many shocks were required or through how much of the training period shocks

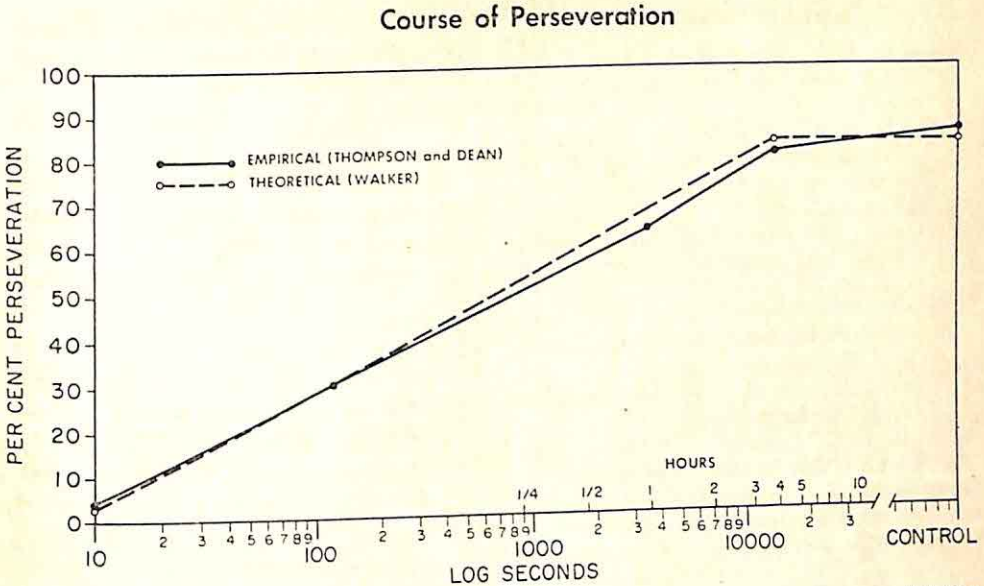


Fig. 9. A comparison of a theoretical prediction of the course of perseverating consolidation with empirical material on the effects of ECS on retention of a discrimination habit. Data taken from Thompson and Dean (12), and replotted here on a log abscissa.

were continued. Finally, Thompson and Dean gave their animals massed training on an easy horizontal-vertical problem. Each animal was trained on this problem in the shortest possible period of time; an average of about 28 minutes seems a likely estimate from their description. One group of such animals was given electroconvulsive shocks 10 seconds after the end of this training, another group was shocked two minutes after training, a third group one hour after training, a fourth group four hours after training, and a control group was not subjected to shocks. Forty-eight hours later they were retrained, and the relative savings in relearning were computed.

The results may be seen in Fig. 9. The dotted line is the theoretical prediction. The solid one is a replot of the error results reported by Thompson and Dean, now placed on a log scale. The apparent closeness is probably adventitious. From the theory, one would expect the duration of the action decrement to vary widely with the relative ease or difficulty of the problems presented, with the type and level of motive, with the magnitude and delay of reward, and with many other factors. The Walker study (14), on which the theoretical curve is predicated, and the study by Thompson and Dean (13) differed on all of these dimensions. Either the period of decrement and perseverative consolidation is much less subject to variation through the manipulation of these variables, or the closeness of fit in Fig. 9 is a most happy pair of combinations of variables.

In summary, it can be argued that there is some merit to the position that the action decrement represents a trace of the perseverative consolidation process, that it is thus positively related to learning, at least in some instances, and that one can expect operations which produce rapid learning to produce great decrements. Wherever it is applicable, this theory is in complete contradiction to the position that the fatigue-like decrement is detrimental to or antithetical to habit strength and learning.

REFERENCES

1. ECCLES, J. C. *The neurophysiological basis of mind.* Oxford: Clarendon Press, 1953.
2. ECCLES, J. C., & RALL, W. Effects induced in a monosynaptic reflex path by its activation. *J. Neurophysiol.,* 1951, **14**, 353–376.
3. GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.,* 1953, **45**, 387–393.
4. HEBB, D. O. Drives and the CNS (conceptual nervous system). *Psychol. Rev.,* 1955, **62**, 243–254.
5. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
6. HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1950.
7. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
8. KAMIN, L. J. Traumatic avoidance learning: The effects of CS-US interval with a trace-conditioning procedure. *J. comp. physiol. Psychol.,* 1954, **47**, 65–72.
9. MONTGOMERY, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.,* 1952, **45**, 50–57.
10. OLDS, J. Physiological mechanisms of reward. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1955. Pp. 73–138.
11. OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation in the septal area and other regions of the rat brain. *J. comp. physiol. Psychol.,* 1954, **47**, 419–427.
12. SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain,* 1956, **79**, 655–680.
13. THOMPSON, R., & DEAN, W. A further study on the retroactive effects of ECS. *J. comp. physiol. Psychol.,* 1955, **48**, 488–491.
14. WALKER, E. L. The course and duration of the reaction decrement and the influence of reward. *J. comp. physiol. Psychol.,* 1956, **49**, 167–176.
15. WALKER, E. L., DEMBER, W. N., EARL, R. W., & KAROLY, A. J. Choice alternation: I. Stimulus vs. place vs. response. *J. comp. physiol. Psychol.,* 1955, **48**, 19–23.
16. WALKER, E. L., DEMBER, W. N., EARL, R. W., FAWL, C. L., & KAROLY, A. J. Choice alternation: III. Response intensity vs. response discriminability. *J. comp. physiol. Psychol.,* 1955, **48**, 80–85.
17. WALKER, E. L., & PARADISE, N. E. An experimental test of the positive correlation between action decrement and learning. *J. exp. Psychol.,* in press.

(Received August 12, 1957)

Psychological Review
Vol. 65, No. 3, 1958

NONCONSTRUCTURAL ASPECTS OF PSYCHOLOGICAL CONSTRUCTS

W. W. MEISSNER

Saint Louis University

Much of what has been written on the subject of psychological theory construction, both on the part of psychological theorists and on the part of the philosophers of science, has been directed to the analysis and determination of the nature and function of the concepts which are found in such theoretical systems. The emphasis seems to have centered on the refinement of the methodology by which such concepts are formed, and on the meaning status of the constructs as they function in the system. Thus, operationism and the constructural method have received extensive treatment. They provide the method and the meaning for a scientific psychology.

Careful examination of the current psychological systems will reveal certain conceptual elements which function in the formulation and application of these theories, but which do not fall under the aegis of the formally defined and accepted methods by which the systems have been constructed. This paper will attempt to indicate some of these nonconstructural aspects, and briefly discuss one or two of the implications that flow from them.

Obviously, there are nonsystematic, nonscientific, nonconstructural operations and processes that enter into the construction of a theoretical system by way of necessary preliminaries. The scientist must read instruments, take measurements, feed rats, and a thousand other things, which are not directly incorporated into his theory, but which provide the data from which he will ultimately construct his system. We are not concerned with these pretheoretical elements. But we are concerned with other nonconstructural aspects which reveal themselves in the very construction and application of the theoretical constructs themselves, and may even alter the meaning of the theoretical concepts.

Our investigation will focus on three such nonconstructural elements: the attribution of real existence (as a real part of the organism) to theoretical constructs, the theoretical assumption of the construct as a quasi-real structure for purely methodological purposes (without asserting its real existence in the organism), and lastly, the experiential reinterpretation of theoretical concepts which have been previously defined formally by systematic methods.

Since the appearance of Bridgman's introduction of the concept of the operational definition into the methodology of psychology, psychological theorists have been deeply concerned over the adequacy of operationally defined constructs to function as descriptions of psychological phenomena (12). Bridgman's explicitation of the operational method focused attention on the character of the constructions which were related to any given complex of operations. Tolman (32, 33) identified and used the intervening variable as a type of construct mediating operationally defined variables. Although he seems to have intended the intervening variable as a purely functional and relational concept, it is not clear that this purely relational concept did not turn out to be something quite different (31).

The effort of MacCorquodale and Meehl (19) to clarify the terminology applied to these constructions was a start in the direction of identifying the kinds of concepts which could function in a psychological theory. They distinguished the intervening variable from the hypothetical construct. The intervening variable was an abstract, empirical relation, which merely served to organize lower level empirical laws. The hypothetical construct, however, involved some supposition of entities or processes, which were not directly observed, and therefore were not included in the original definitions. Essentially, this "surplus meaning" was taken to be the implication of the existence of entities and the occurrence of events in the real subject of investigation. However, it seems that this interpretation of hypothetical constructs and their "surplus meaning" is rather limited and has led to a certain amount of unnecessary confusion.

Supposing for the moment that "surplus meaning" has reference primarily to the suggestion of existence (it may signify other things),then what interpretations can be put on this "existence"? In the original sense of real existence, as suggested by MacCorquodale and Meehl, there is necessary a further distinction. "Real existence" of an inferred entity can be taken as meaning that the inferred entity, as formally understood, really exists as part of the intrinsic, objective, noninferred, real structure, which constitutes the object of investigation. Or, the "real existence" of an inferred entity can be understood to mean only that the construct is formed and intended as a theoretical correlative of the real structure of the object. In this sense, a hypothetical construct is said to denote (10). Lindzey (18) has expressed the need for this sort of clarification in his attempt at the description of what he termed "conventional constructs." This type of construct involves an inference about the reality of the organism no less that any hypothetical construct—but it is precisely an inference about the organism *as it is known in the theory*. It is not an inference about the real structure of the organism itself. The former sense, if found in the formulation and interpretation of psychological theories, constitutes a gross misunderstanding of the nature of theoretical constructions, as well as a misapplication of methodology. The reasons for this will be discussed below. The latter sense, however, is a legitimate systematic concept, and is the only acceptable and possible meaning of "real existence" as applied to hypothetical constructs. Since this sort of "surplus meaning" is not found in the constructed definition of the concept, however, it is a nonconstructural element within the framework of the theory.

This theoretical-real existence must not be confused with the kind of existence which is applied to a construct by reason of its contextual consistency within the framework of its proper theory. This sort of "existence" is nothing but the coherence of logical relations among the constructural concepts which compose the theory. The logical coherence of constructs in a theory can not be taken as indicative of anything beyond the theory itself. This sort of *systemic existence* (1) is common to both intervening variables and hypothetical constructs. The "surplus meaning" involving "real existence" of hypothetical constructs must be reserved to theoretical-real existence in the latter sense of a theoretical correlate described above.

The failure to recognize this distinction between meanings of real existence has led to a reconfusion of terminology (25). The supposition of theoretical-real existence has sometimes been assimilated to the intervening variable (22, 23), with the result that the intervening variable is no longer merely intervening, and the hypothetical construct is reduced to the level of a preliminary and imperfect stage in the development of a fully operational theory based on the intervening-variable type of construct alone. If Feigl's "existential assumption" can be interpreted in this sense of theoretical-real existence, it would seem that this plea for the return of the hypothetical construct is fundamentally an appeal to this sort of distinction in the meaning of "real existence" (7, 8, 9).

Correlative to the issue of real existence is the question of the conceptual status of constructs and constructural models as they function in theory. The issue of the legitimacy or illegitimacy of the flat assertion that theoretical entities have real existence will be decided in terms of our understanding of the nature of these concepts. There is no question of the legitimacy of constructural concepts, but only of the assertion of their existence. Generally speaking, methodologists are aware that a theoretical model can not be understood as representative of the real structure of some real entity (28, 31). This is fairly obvious in the case of the intervening variable, whose characteristics are of a purely formal nature. But in the case of the hypothetical construct, where some sort of objectivized (not objective) structure is implied, the issue is raised again and again.

A correct appraisal of the nature of constructural methodology has led many theorists to evaluate the legitimate "surplus meaning" (i.e., theoretical-real existence) of the hypothetical construct as an extrasystematic and, therefore, nonconstructural element, which introduces an impure fraction into the constructural purity of the theory. However, the surplus enters into the understanding of such a construct only to the extent that the construct itself is structural (15). The hypothetical construct differs from an intervening variable precisely in this, that it hypothesizes some sort of structure to explain the relation of lower-level variables (25). Obviously, a scientist can not think of structure, as opposed to a merely formal or mathematical function, without thinking of it in terms of "theoretical-real existence." This does not imply anything about the *real* structure of any *real* thing, but it does imply that the theorist is thinking of this construct as an analogue to the real structure of a real thing.

Theoretical models must be thought of then as research tools, which the theorist forms without direct dependence on or inference from the immediate data of experience. The constructural-theoretical model is related to the ultimate real structure of the thing under investigation through a complex set of relations, which involve certain specified methodological procedures, selective processes, and measurement techniques. The point here is that theoretical constructs are not directly tied down or determined by the real structure of their objects. If they were so tied to the structure of the thing, there would be little room for the multiple hypothetical models we find in psychology, nor could the theorist enjoy the relative freedom in structuring the model that is now his. It is for this reason that the assertion of the noninferred real existence of such theoretical entities must be looked upon as illegitimate.

Models should not really be thought of as primarily imitative, but rather as a sort of adaptable pattern, which can be compared with the available data of experiment to find out what adaptations are to be made and in what direction the course of our investigation must be turned (21). Although it is not clear to me what the criteria of usefulness are to be in the development of a psychological theory, the recent emphasis on "usefulness" of theoretical constructs, at least for the formulation of research hypotheses, marks an encouraging shift from the former preoccupation with reality (30). Even on logical grounds, it has been pointed out (20) that the common procedure of testing theories by inference and confirmation is open to the possible fallacy that the consequence is unique for that theory. If Consequence B is deduced from Theory A and experimentally confirmed, we have added to the probability of Theory A; but we cannot assume that the same Consequence B may not be implied by some alternative theory. The hypothetical construct is a surrogate or symbol which is constructed in the theorist's mind as a substitute for the real structure of the organism, and which is only indirectly related to the structure of that organism. The whole direction or impulse of theory construction is to the progressive formulation of better theoretical constructions which are capable of more rigorous evaluation and which will gradually converge on the real structure of the organism (24).

Feigl (6) has expressed the belief that higher-order hypothetical constructs, which are built on and constructed from operations, will gain real insight into the natures of things. But in a later writing (8), he recognizes that the meaning of a psychological concept is generally not identical with its defining operations. If we are able to speak of insight into the natures of things (whatever that may mean in the context of constructural theory), whatever meaning is contained in higher-level constructs must be reflected in the lower-level definitions in terms of observable data. If meaning and definition are not identical, we must explain the difference. I do not mean that we must explain the fact of a difference, but we must define the difference. Otherwise our procedure is unscientific and methodologically unsound.

Along this same line, Krech (17) has pointed out that hypothetical constructs could not function in a purely psychological psychology, since the purely psychological data are beyond the reach of scientific inquiry. Because the experience of a private datum is an unverifiable event, and because science is concerned only with verifiable events, some would claim that a scientific psychology of private experience is a contradiction (32). Either psychology proceeds behavioristically and is a science, or it treats of the private datum and relinquishes its claim to scientific status (4). Thus Krech has indicated that the concepts of tensions, needs, and cognitive structures are sometimes defined as hypothetical constructs, and sometimes as conscious experience or behavior (17). This means that there has been a shift in the formal meaning of these concepts, and we are dealing with two concepts instead of only one. One is a formally defined hypothetical construct, which functions as an integral part of a theory. As such, it is linked to the observable data, from which it is structured, by low-level empirical laws and operational definitions. The other concept, which represents the redefinition of the previous hypothetical construct in terms of experience, is more difficult to delineate because of the fact that it is rarely, if ever, given any sort of explicit or formal presentation. However, in general, it seems to take the form of a subtle reinterpretation of a previously defined concept in terms of some sort of phenomenological experience. This sort of shift in meaning has been pointed out as due to a failure to distinguish between the operational meaning of intervening variables or hypothetical constructs and the intuitive properties ascribed to them (14).

This shift in the interpretation of concepts must not be confused with the possible shift that sometimes occurs in the construction of theoretical systems between intervening variables and hypothetical constructs. It has sometimes happened that a concept is defined and constructed as a purely functional intervening variable, but later in the development of the theory it acquires a denotative meaning and a certain amount of structural connotation, which qualifies it as a hypothetical construct (10). Both Tolman (26) and Hull (10, 19, 25, 27) have been accused of permitting this subtle reorientation in the meaning of their concepts. Very likely, Bergmann's dismissal of the distinction between intervening variable and hypothetical construct as a pseudodistinction (3) stems from insufficient attention to the possibility of this sort of subtle shift in the formal meaning of concepts.

As long as the surplus meaning, which constitutes the concept as a hypothetical construct, involves no more than the assumption of theoretical-real existence, the procedure is by no means illegitimate, and can be scientifically fruitful. It is this sense, I believe, that Feigl has spoken of "existential assumptions" (7, 8). However, if the surplus meaning involves the implication of real existence (involving attribution of the hypothetical construct to the organism as part of the real structure of the organism), the connotation is illegitimate and unscientific, as we have already pointed out. There seems to be a certain amount of evidence to suggest that Hull was not always able to avoid this error. In regard to the *sHr*, he wrote that "it exists as an organization as yet largely unknown, hidden within the complex structure of the nervous system" (11, p. 102). Likewise, MacCorquodale and Meehl were quick to suggest that Hull's early articles on conditioning were written because Hull "considered that the hypothetical events represented in his diagrams may have actually *occurred* and that the occurrence of the events represents the underlying truth about the learning phenomena he dealt with" (19, pp. 104–105).

In our discussion thus far, we have seen that this sort of implication of the real existence of an inferred entity as a part of the real structure of an organism is extrasystematic and to that extent, nonconstructural. This also constitutes an illegitimate inference, which is not only not supported by the defining method, but is contrary to the very nature of the scientific constructural methodology. We have also seen that the theoretical assumption of the constructed concept as correlative, and perhaps suggestive of the real structure of the organism, is again extrasystematic and nonconstructural, but in this case the assumption of existence (which is equivalently an assumption of *non*existence) is entirely legitimate, and in fact a necessary correlate of the development of theoretical concepts. We have further suggested a third kind of nonconstructural ingredient which seems to function in psychological theory: the reinterpretation of systematically constructed concepts in terms of experiential variables. This last nonconstructural aspect deserves further consideration.

Spence (31) has already pointed out that the method of field theorists like Lewin depends heavily on phenomenological introspection for the introduction of theoretical concepts. If these concepts are then defined systematically in terms of given observables, they function as integral parts of a scientifically constructed theory. But, if they continue to be applied and interpreted in terms of the original experiential context in which they arose, they constitute something else by way of formal concepts. It is quite possible that the correlation of empirical constructs with formal terms according to the logic of dynamics, as Koch (16) has analyzed them in Lewin's system, presents a difficulty precisely in so far as the empirical constructs are functioning at a phenomenological level and the formal terms are presented at a constructural level.

An example may help to clarify the problem. Hull (11) introduces his concept of habit strength as a functional stimulus-response relation between stimulus energy, response, nature and number of reinforcements, and the reaction potential, which is measured in terms of the actual response. This intervening-variable type of concept is then conceived as a dynamic relationship between the afferent and efferent neural impulses. This seems to be more of a hypothetical construct, but the definition of the term is still couched in systematic and operational terms, and the switch from intervening variable to hypothetical construct is still acceptable. However, at one point Hull tells us that "it would appear that the presence of internal conditions or neural organizations, such as habit structures (*sHr*) and reaction potentials (*sEr*) are also reportable" by introspection (13, p. 344). He is also explicit in stating his conception of receptor-effector connections in terms of common sense habits (11). If Hull is able to identify these constructs by introspective methods, it follows that the concept, as operationally defined and identified, is now being reinterpreted and reidentified in terms of experiential elements which are not included in the systematic formulation of these concepts. Phenomenological introspection is not restricted to field theorists alone, so it seems.

The introduction of these nonconstructural elements into even the most sophisticated of behavioral systems raises a serious question. Snygg (29) has raised the issue of a phenomenological psychology. But, like Tolman (32) before him, he draws the line between the phenomenological and the objective in such a way as to give us a none too happy choice. Are we forced to make such a choice? The scientific method, as it has been transferred from natural sciences to psychology, seems to have exhausted a good deal of its potentiality (5). Further, as a method, it is not equipped to deal with purely psychological data. As a result, and this is the point I have tried to make in this paper, psychological theory has resorted to a subtle, almost unconscious, translation of its systematic concepts in terms of psychological experience. This step has been necessary, in a sense, in order that psychological constructs assume some semblance of pertinence and meaningfulness for the human psychological situation. And this experiential interpolation is a necessary development in all psychological theory, for it is only in terms of this sort of interpolation that psychological theory can function as significant for the problems of applied human psychology, and for the prediction and control of behavior.

Rather than interpret this analysis as disjunctive and exclusive—and this seems to be the prevailing attitude of psychological theorists—it might prove promising to take it as conjunctive and inclusive. I do not mean that the experiential and objective aspects are reducible to common grounds. Nor do I mean that one or the other should be eliminated. The objective, scientific techniques developed by experimental psychology are invaluable, and any attempt at devaluating or reducing them would be unrealistic, to say the least. Experimental methods are a necessary means of introducing control, confirmation, and validation into theory construction. But it does not seem that these methods are in themselves sufficient to meet the demands of a scientific psychology. Our theories are now subject to experiential reinterpretation which takes on the color of a necessary adjunct to their significance and applicability. As the situation now stands, this experiential interpolation is unhealthy from the standpoint of scientific theory because it is extrasystematic and has no formal place in our theory construction. For this reason, the more rigorous theorists have tended to reject it explicitly, although they cannot avoid it implicitly (2).

But, it is not entirely clear that phenomenological or experiential aspects need be so excluded. If these elements can be made explicit in our theories and can be justified in our methodologies, the present status of psychology would be much healthier than it is. The formulation of a distinctively psychological methodology, based on the inclusion of both objective and experiential methods, and giving rise to theoretical construction of a distinctively psychological character, is a very tempting goal. The reconstruction of such a distinctive and comprehensive methodology must be left to the future. But at least a first step in its direction must be taken by pointing out the revelant elements as they function in the theories we have at the present time.

SUMMARY

Three nonconstructural elements, which appear in current psychological constructural systems, are singled out and briefly discussed: (*a*) attribution to a theoretical construct of "real existence," as if the construct were actually a description of the real structure of the organism, rather than of the organism *as it is known in the theory;* (*b*) attribution of "theoretical-real existence" to a construct, without attributing the defined structure to the real organism which is merely a methodological assumption and implies nothing about the real organisms; (*c*) phenomenological or experiential reinterpretation of theoretical constructs. Of these nonconstructural aspects, the first is illegitimate, the second is legitimate and scientifically sound, and the last is discussed in the light of possible evolution of a distinctively psychological methodology which would incorporate both phenomenological and behavioristic elements.

REFERENCES

1. BECK, L. W. Constructions and inferred entities. *Phil. Sci.*, 1950, **17**, 74–86.
2. BERGMANN, G. On some methodological problems of psychology. *Phil. Sci.*, 1940, **7**, 205–219.
3. BERGMANN, G. Theoretical psychology. *Ann. Rev. Psychol.*, 1953, **4**, 435–458.
4. CONWAY, B. J. W. Philosophy, metaphysics and psychology. In *Proceedings of the Second International Congress of the International Union for the Philosophy of Science, 1954*. Vol. V. Zurich: Neuchatel Suisse, 1955. Pp. 79–80.
5. FARRELL, B. A. On the limits of experimental psychology. *Brit. J. Psychol.*, 1955, **46**, 165–177.
6. FEIGL, H. Rejoinders and second thoughts (Symposium on operationism). *Psychol. Rev.*, 1945, **52**, 284–288.
7. FEIGL, H. Existential hypotheses. *Phil. Sci.*, 1950, **17**, 35–62.
8. FEIGL, H. Principles and problems of theory construction in psychology. In W. Dennis (Ed.), *Current trends in psychological theory*. Pittsburgh: Univer. Pittsburgh Press, 1951. Pp. 179–213.
9. FEIGL, H. Functionalism, psychological theory, and the uniting sciences: Some discussion remarks. *Psychol. Rev.*, 1955, **62**, 232–235.
10. GINSBERG, A. Hypothetical constructs and intervening variables. *Psychol. Rev.*, 1954, **61**, 119–131.
11. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
12. HULL, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291.
13. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
14. KENDLER, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.*, 1952, **59**, 269–277.
15. KESSEN, W., & KIMBLE, G. A. "Dynamic systems" and theory construction. *Psychol. Rev.*, 1952, **59**, 263–267.

16. KOCH, S. The logical character of the motivation concept. *Psychol. Rev.,* 1941, **48**, 15–37; 127–154.

17. KRECH, D. Dynamic systems, psychological fields, and hypothetical constructs. *Psychol. Rev.,* 1950, **57**, 283–290.

18. LINDZEY, G. Hypothetical constructs, conventional constructs, and the use of physiological data in psychological theory. *Psychiatry,* 1953, **16**, 27–33.

19. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.,* 1948, **55**, 95–107.

20. MCGUIGAN, F. J. Confirmation of theories in psychology. *Psychol. Rev.,* 1956, **63**, 98–103.

21. MACKAY, D. M. On comparing the brain with machines. *Amer. Scientist,* 1954, **42**, 261–268.

22. MARX, M. H. The general nature of theory construction. In M. H. Marx (Ed.), *Psychological theory.* New York: MacMillan, 1951. Pp. 4–19.

23. MARX, M. H. Intervening variable or hypothetical construct? *Psychol. Rev.,* 1951, **58**, 235–247.

24. MARZOCCO, F. N. The character of psychological theories. *J. gen. Psychol.,* 1955, **52**, 197–203.

25. MAZE, J. R. Do intervening variables intervene? *Psychol. Rev.,* 1954, **61**, 226–234.

26. O'NEIL, W. M. Hypothetical terms and relations in psychological theorising. *Brit. J. Psychol.,* 1953, **44**, 211–220.

27. ROZEBOOM, W. W. Mediation variables in scientific theory. *Psychol. Rev.,* 1956, **63**, 249–264.

28. SHOBEN, E. J. Psychological theory construction and the psychologist. *J. gen. Psychol.,* 1955, **52**, 181–188.

29. SNYGG, D. The need for a phenomenological system of psychology. *Psychol. Rev.,* 1941, **48**, 404–424.

30. SNYGG, D. Scientific method in psychology. *J. gen. Psychol.,* 1955, **52**, 189–196.

31. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.,* 1944, **51**, 47–68.

32. TOLMAN, E. C. Psychology versus immediate experience. *Phil. Sci.,* 1935, **2**, 356–380.

33. TOLMAN, E. C. Operational behaviorism and current trends in psychology. In *Proc. 25th Ann. Celebr. Inaug. Grad. Stud.* Los Angeles: Univer. Southern California Press, 1936. Pp. 89–103.

(Received May 31, 1957)

Psychological Review
Vol. 65, No. 3, 1958

ELEMENTS OF A THEORY OF HUMAN PROBLEM SOLVING

ALLEN NEWELL, J. C. SHAW
The RAND Corporation
AND HERBERT A. SIMON
Carnegie Institute of Technology

In this paper we shall set forth the elements of a theory of human problem solving, together with some evidence for its validity drawn from the currently accepted facts about the nature of problem solving. What questions should a theory of problem solving answer? First, it should predict the performance of a problem solver handling specified tasks. It should explain how human problem solving takes place: what processes are used, and what mechanisms perform these processes. It should predict the incidental phenomena that accompany problem solving, and the relation of these to the problem-solving process. For example, it should account for "set" and for the apparent discontinuities that are sometimes called "insight." It should show how changes in the attendant conditions—both changes "inside" the problem solver and changes in the task confronting him—alter problem-solving behavior. It should explain how specific and general problem-solving skills are learned, and what it is that the problem solver "has" when he has learned them.

Information Processing Systems

Questions about problem-solving behavior can be answered at various levels and in varying degrees of detail. The theory to be described here explains problem-solving behavior in terms of what we shall call *information processes*. If one considers the organism to consist of effectors, receptors, and a control system for joining these, then this theory is mostly a theory of the control system. It avoids most questions of sensory and motor activities. The theory postulates:

1. A control system consisting of a number of *memories*, which contain symbolized information and are interconnected by various ordering relations. The theory is not at all concerned with the physical structures that allow this symbolization, nor with any properties of the memories and symbols other than those it explicitly states.

2. A number of *primitive information processes*, which operate on the information in the memories. Each primitive process is a perfectly definite operation for which known physical mechanisms exist. (The mechanisms are not necessarily known to exist in the human brain, however—we are only concerned that the processes be described without ambiguity.)

3. A perfectly definite set of rules for combining these processes into whole *programs* of processing. From a program it is possible to deduce unequivocally what externally observable behaviors will be generated.

At this level of theorizing, *an explanation of an observed behavior of the organism is provided by a program of primitive information processes that generates this behavior.*

A program viewed as a theory of behavior is highly specific: it describes one organism in a particular class of situations. When either the situation or the organism is changed, the program must be modified. The program can be used as a theory—that is, as a predictor of behavior—in two distinct ways. First, it makes many precise predictions that can be tested in detail regarding the area of behavior it is designed to handle. For example, the theory considered in this paper predicts exactly how much difficulty an organism with the specified program will encounter in solving each of a series of mathematical problems: which of the problems it will solve, how much time (up to a proportionality constant) will be spent on each, and so on.

Second, there will be important qualitative similarities among the programs that an organism uses in various situations, and among the programs used by different organisms in a given situation. The program that a human subject uses to solve mathematical problems will be similar in many respects to the program he uses to choose a move in chess; the program one subject uses for any such task will resemble the programs used by other subjects possessing similar training and abilities. If there were no such similarities, if each subject and each task were completely idiosyncratic, there could be no theory of human problem solving. Moreover, there is some positive evidence, as we shall see, that such similarities and general characteristics of problem-solving processes do exist.

In this paper we shall limit ourselves to this second kind of validation of our theory of problem solving. We shall predict qualitative characteristics of human problem-solving behavior and compare them with those that have already been observed and described. Since all of the available data on the psychology of human problem solving are of this qualitative kind, no more detailed test of a program is possible at present. The more precise validation must wait upon new experimental work.[1]

In succeeding sections we shall describe an information-processing program for discovering proofs for theorems in logic. We shall compare its behavior qualitatively with that of human problem solvers. In general, the processes that compose the program are familiar from everyday experience and from research on human problem solving: searching for possible solutions, generating these possibilities out of other elements, and evaluating partial solutions and cues. From this standpoint there is nothing particularly novel about the theory. It rests its claims on other considerations:

1. It shows specifically and in detail how the processes that occur in human problem solving can be compounded out of elementary information processes, and hence how they can be carried out by mechanisms.

2. It shows that a program incorporating such processes, with appropriate organization, can in fact solve problems. This aspect of problem solving has been thought to be "mysterious" and unexplained because it was not understood how sequences of simple processes could account for the successful solution of complex problems. The theory dissolves the mystery by showing that nothing more need be added to the constitution of a successful problem solver.

Relation to Digital Computers

The ability to specify programs precisely, and to infer accurately the behavior they will produce, derives from the use of high-speed digital computers. Each specific theory—each program of information processes that purports to describe some human behavior—is coded for a computer. That is, each primitive information process is coded to be a separate computer routine, and a "master" routine is written that allows these primitive processes to be assembled into any system we wish to specify. Once this has been done, we can find out exactly what behavior the purported theory predicts by having the computer "simulate" the system.

---

[1] Several studies of individual and group problem-solving behavior with logic problems have been carried out by O. K. Moore and Scarvia Anderson (5). The problems Moore and Anderson gave their subjects are somewhat different from those handled by our program, and hence a detailed comparison of behavior is not yet possible. We are now engaged, with Peter Houts, in replicating and extending the experiments of Moore and Anderson with human subjects and at the same time modifying our program to predict the human laboratory behavior in detail.

We wish to emphasize that we are not using the computer as a crude analogy to human behavior—we are not comparing computer structures with brains, nor electrical relays with synapses. Our position is that the appropriate way to describe a piece of problem-solving behavior is in terms of a program: a specification of what the organism will do under varying environmental circumstances in terms of certain elementary information processes it is capable of performing. This assertion has nothing to do—directly—with computers. Such programs could be written (now that we have discovered how to do it) if computers had never existed.[2] A program is no more, and no less, an analogy to the behavior of an organism than is a differential equation to the behavior of the electrical circuit it describes. Digital computers come into the picture only because they can, by appropriate programming, be induced to execute the same sequences of information processes that humans execute when they are solving problems. Hence, as we shall see, these programs describe both human and machine problem solving at the level of information processes.[3]

With this discussion of the relation of programs to machines and humans behind us, we can afford to relax into convenient, and even metaphoric, uses of language without much danger of misunderstanding. It is often convenient to talk about the behavior implied by a program as that of an existing physical mechanism doing things. This mode of expression is legitimate, for if we take the trouble to put any particular program in a computer, we have in fact a machine that behaves in the way prescribed by the program. Similarly, for concreteness, we will often talk as if our theory of problem solving consisted of statements about the ability of a computer to do certain things.

THE LOGIC THEORIST

We can now turn to an example of the theory. This is a program capable of solving problems in a particular domain—capable, specifically, of discovering proofs for theorems in elementary symbolic logic. We shall call this program the Logic Theorist (LT).[4] We assert that the behavior of this program, when the stimulus consists of the instruction that it prove a particular theorem, can be used to predict the behavior of (certain) humans when they are faced with the same problem in symbolic logic.

---

[2] We can, in fact, find a number of attempts in the psychological literature to explain behavior in terms of programs—or the prototypes thereof. One of the most interesting, because it comes relatively close to the modern conception of a computer program, is Adrian de Groot's analysis of problem solving by chess players (2). The theory of de Groot is based on the thought-psychology of Selz, a somewhat neglected successor to the Wurzburg school. Quite recently, and apparently independently, we find the same idea applied by Jerome S. Bruner and his associates to the theory of concept formation (1). Bruner uses the term "strategy," derived from economics and game theory, for what we have called a program.

[3] For a fuller discussion of this point see (9).

[4] In fact, matters are a little more complicated, for in the body of this paper we will consider both the basic program of LT and a number of variants on this program. We will refer to all of these variants, interchangeably, as "LT." This will not be confusing, since the exact content of the program we are considering at any particular point will always be clear from the context.

The program of LT was not fashioned directly as a theory of human behavior; it was constructed in order to get a program that would prove theorems in logic. To be sure, in constructing it the authors were guided by a firm belief that a practicable program could be constructed only if it used many of the processes that humans use. The fact remains that the program was not devised by fitting it directly to human data. As a result, there are many details of LT that we would not expect to correspond to human behavior. For example, no particular care was exercised in choosing the primitive information processes to correspond, point by point, with elementary human processes. All that was required in writing the program was that the primitive processes constitute a sufficient set and a convenient set for the type of program under study.

Since LT has been described in detail elsewhere (6, 8), the description will not be repeated here. It will also be unnecessary to describe in detail the system of symbolic logic that is used by LT. For those readers who are not familiar with symbolic logic, we may remark that problems in the sentential calculus are at about the same level of difficulty and have somewhat the same "flavor" as problems in high school geometry.[5]

Design of the Experiments

First we will describe the overt behavior of LT when it is presented with problems in elementary symbolic logic. In order to be concrete, we will refer to an experiment conducted on a digital computer. We take an ordinary general-purpose digital computer,[6] and store in its memory a program for interpreting the specifications of LT. Then we load the program that specifies LT. The reader may think of this program as a collection of techniques that LT has acquired for discovering proofs. These techniques range from the ability to read and write expressions in symbolic logic to general schemes for how a proof might be found.

Once we have loaded this program and pushed the start button, the computer, to all intents and purposes, *is* LT. It already knows how to do symbolic logic, in the sense that the basic rules of operation of the mathematics are already in the program (analogously to a human's knowing that "equals added to equals give equals" in elementary algebra).

We are now ready to give LT a task. We give it a list of the expressions (axioms and previously proved theorems) that it may take as "given" for the task at hand. These are stored in LT's memory. Finally, we present LT with another expression and instruct it to discover a proof for this expression.

From this point, the computer is on its own. The program plus the task uniquely determines its behavior. It attempts to find a proof—that is, it tries various techniques, and if they don't work, it tries other techniques. If LT finds a legitimate proof, it prints this out on a long strip of paper. There is, of course, no guarantee that it will find a proof; after working for some time, the machine will give up—that is, it will stop looking for a proof.

---

[5] LT employs the sentential calculus as set forth in Chapters 1 and 2 of A. N. Whitehead and Bertrand Russell, *Principia Mathematica* (10)—the "classic" of modern symbolic logic. A simple introduction to the system of *Principia* will be found in (3).

[6] The experiments described here were carried out with the RAND JOHNNIAC computer. The JOHNNIAC is an automatic digital computer of the Princeton type. It has a word length of 40 bits, with two instructions in each word. Its fast storage consists of 4,096 words of magnetic cores, and its secondary storage consists of 9,216 words on magnetic drums. Its speed is about 15,000 operations per second. The programming techniques used are described more fully in (6). The experiments are reported in more detail in (7).

Now the experimenters know exactly what is in the memory of LT when it starts—indeed, they created the program. This, however, is quite different from saying that the experimenters can predict everything LT will do. In principle this is possible; but in fact the program is so complex that the only way to make detailed predictions is to employ a human to simulate the program by hand. (A human can do anything a digital computer can do, although it may take him considerably longer.)

1. As the initial experiment, we stored the axioms of *Principia Mathematica*, together with the program, in the memory of LT, and then presented to LT the first 52 theorems in Chapter 2 of *Principia* in the sequence in which they appear there. LT's program specified that as a theorem was proved it was stored in memory and was available, along with the axioms, as material for the construction of proofs of subsequent theorems. With this program and this order of presentation of problems, LT succeeded in proving 38 (73%) of the 52 theorems. About half of the proofs were accomplished in less than a minute each; most of the remainder took from one to five minutes. A few theorems were proved in times ranging from 15 minutes to 45 minutes. There was a strong relation between the times and the lengths of the proofs—the time increasing sharply (perhaps exponentially) with each additional proof step.

2. The initial conditions were now restored by removing from LT's memory the theorems it had proved. (Translate: "A new subject was obtained who knew how to solve problems in logic but was unfamiliar with the particular problems to be used in the experiment.") When one of the later theorems of Chapter 2 (Theorem 2.12) was presented to LT, it was not able to find a proof, although when it had held the prior theorems in memory, it had found one in about ten seconds.

3. Next, an experiment was performed intermediate between the first two. The axioms and Theorem 2.03 were stored in memory, but not the other theorems prior to Theorem 2.12, and LT was again given the task of proving the latter. Now, using Theorem 2.03 as one of its resources, LT succeeded—in fifteen minutes—where it had failed in the second experiment. The proof required three steps. In the first experiment, with all prior theorems available, the proof required only one step.

Outcome of the Experiments

From these three series of experiments we obtain several important pieces of evidence that the program of LT is qualitatively like that of a human faced with the same task. The first, and most important, evidence is that LT does in fact succeed in finding proofs for a large number of theorems.

Let us make this point quite clear. Since LT can actually discover proofs for theorems, its program incorporates a *sufficient* set of elementary processes arranged in a sufficiently effective strategy to produce this result. Since no other program has ever been specified for handling successfully these kinds of problem-solving tasks, no definite alternative hypothesis is available. We are well aware of the standard argument that "similarity of function does not imply similarity of process." However useful a caution this may be, it should not blind us to the fact that specification of a set of mechanisms sufficient to produce observed behavior is strong confirmatory evidence for the theory embodying these mechanisms, especially when it is contrasted with theories that cannot establish their sufficiency.

The only alternative problem-solving mechanisms that have been completely specified for these kinds of tasks are simple algorithms that carry out exhaustive searches of all possibilities, substituting "brute force" for the selective search of LT. Even with the speeds available to digital computers, the principal algorithm we have devised as an alternative to LT would require times of the order of hundreds or even thousands of years to prove theorems that LT proves in a few minutes. LT's success does not depend on the "brute force" use of a computer's speed, but on the use of heuristic processes like those employed by humans.[7] This can be seen directly from examination of the program, but it also shows up repeatedly in all the other behavior exhibited by LT.

The second important fact that emerges from the experiments is that LT's success depends in a very sensitive way upon the order in which problems are presented to it. When the sequence is arranged so that before any particular problem is reached some potentially helpful intermediate results have already been obtained, then the task is easy. It can be made progressively harder by skipping more and more of these intermediate stepping-stones. Moreover, by providing a single "hint," as in the third experiment (that is, "Here is a theorem that might help"), we can induce LT to solve a problem it had previously found insoluble. All of these results are easily reproduced in the laboratory with humans. To compare LT's behavior with that of a human subject, we would first have to train the latter in symbolic logic (this is equivalent to reading the program into LT), but without using the specific theorems of Chapter 2 of *Principia Mathematica* that are to serve as problem material. We would then present problems to the human subject in the same sequence as to LT. For each new sequence we would need naive subjects, since it is difficult to induce a human subject to forget completely theorems he has once learned.

PERFORMANCE PROCESSES IN THE LOGIC THEORIST

We can learn more about LT's approximation to human problem solving by instructing it to print out some of its intermediate results—to work its problems on paper, so to speak. The data thus obtained can be compared with data obtained from a human subject who is asked to use scratch paper as he works a problem, or to think aloud.[8] Specifically, the computer can be instructed to print out a record of the subproblems it works on and the methods it applies, successfully and unsuccessfully, while seeking a solution. We can obtain this information at any level of detail we wish, and make a correspondingly detailed study of LT's processes.

To understand the additional information provided by this "thinking aloud" procedure, we need to describe a little more fully how LT goes about solving problems. This description has two parts: (*a*) specifying what constitutes a proof in symbolic logic; (*b*) describing the methods that LT uses in finding proofs.

---

[7] A quantitative analysis of the power of the heuristics incorporated in LT will be found in (7).

[8] Evidence obtained from a subject who thinks aloud is sometimes compared with evidence obtained by asking the subject to theorize introspectively about his own thought processes. This is misleading. Thinking aloud is just as truly behavior as is circling the correct answer on a paper-and-pencil test. What we infer from it about other *processes* going on inside the subject (or the machine) is, of course, another question. In the case of the machine, the problem is simpler than in the case of the human, for we can determine exactly the correspondence between the internal processes and what the machine prints out.

Nature of a Proof

A proof in symbolic logic (and in other branches of logic and mathematics) is a sequence of statements such that each statement: (*a*) follows from one or more of the others that precede it in the sequence, or (*b*) is an axiom or previously proved theorem.[9] Here "follows" means "follows by the rules of logic."

LT is given four rules of inference:

Substitution. In a true expression (for example, "[p or p] implies p") there may be substituted for any variable a new variable or expression, provided that the substitution is made throughout the original expression. Thus, by substituting p or q for p in the expression "(p or p) implies p," we get: "([p or q] or [p or q]) implies (p or q)" but *not:* "([p or q] or p) implies p."

Replacement. In a true expression a connective ("implies," etc.) may be replaced by its definition in terms of other connectives. Thus "A implies B" is defined to be "not-A or B"; hence the two forms can be used interchangeably.

Detachment. If "A" is a true expression and "A implies B" is a true expression, then B may be written down as a true expression.

Syllogism (Chaining). It is possible to show by two successive applications of detachment that the following is also legitimate: If "a implies b" is a true expression and "b implies c" is a true expression, then "a implies c" is also a true expression.

Proof Methods

The task of LT is to construct a proof sequence deriving a problem expression from the axioms and the previously proved theorems by the rules of inference listed above. But the rules of inference, like the rules of any mathematical system or any game, are permissive, not mandatory. That is, they state what sequences *may* legitimately be constructed, not what particular sequence should be constructed in order to achieve a particular result (i.e., to prove a particular problem expression). The set of "legal" sequences is exceedingly large, and to try to find a suitable sequence by trial and error alone would almost always use up the available time or memory before it would exhaust the set of legal sequences.[10]

To discover proofs, LT uses *methods* which are particular combinations of information processes that result in coordinated activity aimed at progress in a particular direction. LT has four methods (it could have more): *substi-*

---

[9] The axioms of symbolic logic and the theories that follow from them are all tautologies, true by virtue of the definitions of their terms. It is their tautological character that gives laws of logic their validity, independent of empirical evidence, as rules of inductive inference. Hence the very simple axioms that we shall use as examples here will have an appearance of redundancy, if not triviality. For example, the first axiom of *Principia* states, in effect, that "if any particular sentence (call it p) is true, or if that same sentence (p) is true, then that sentence (p) is, indeed, true"—for example, "if frogs are fish, or if frogs are fish, then frogs are fish." The "if---then" is trivially and tautologically true irrespective of whether p is true, for in truth frogs are not fish. Since our interest here is in problem solving, not in logic, the reader can regard LT's task as one of manipulating symbols to produce desired expressions, and he can ignore the material interpretations of these symbols.

[10] See (7). The situation here is like that in chess or checkers where the player knows what moves are legal but has to find in a reasonable time a move that is also "suitable"—that is, conducive to winning the game.

tution, detachment, forward chaining, and *backward chaining.* Each method focuses on a single possibility for achieving a link in a proof.

The substitution method attempts to prove an expression by generating it from a known theorem employing substitutions of variables and replacements of connectives.

The detachment method tries to work backward, utilizing the rule of detachment to obtain a new expression whose proof implies the proof of the desired expression. This possibility arises from the fact that if B is to be proved, and we already know a theorem of the form "A implies B," then proof of A is tantamount to proof of B.

Both chaining methods try to work backward to new problems, using the rule of syllogism, analogously to the detachment method. Forward chaining uses the fact that if "a implies c" is desired and "a implies b" is already known, then it is sufficient to prove "b implies c." Backward chaining runs the argument the other way: desiring "a implies c" and knowing "b implies c" yields "a implies b" as a new problem.

The methods are the major organizations of processes in LT, but they are not all of it. There is an executive process that coordinates the use of the methods, and selects the subproblems and theorems upon which the methods operate. The executive process also applies any learning processes that are to be applied. Also, all the methods utilize common subprocesses in carrying out their activity. The two most important subprocesses are the *matching* process, which endeavors to make two given subexpressions identical, and the *similarity test,* which determines (on the basis of certain computed descriptions) whether two expressions are "similar" in a certain sense (for details, cf. 8).

LT can be instructed to list its attempts, successful and unsuccessful, to use these methods, and can list the new subproblems generated at each stage by these attempts. We can make this concrete by an example:

Suppose that the problem is to prove "p implies p." The statement "(p or p) implies p" is an axiom; and "p implies (p or p)" is a theorem that has already been proved and stored in the theorem memory. Following its program, LT first tries to prove "p implies p" by the substitution method, but fails because it can find no similar theorem in which to make substitutions.

Next, it tries the detachment method. Letting B stand for "p implies p," several theorems are found of the form "A implies B." For example, by substitution of not-p for q, "p implies (q or p)" becomes "p implies (not-p or p)"; this becomes, in turn, by replacement of "or" by "implies": "p implies (p implies p)." Discovery of this theorem creates a new subproblem: "Prove A"—that is, "prove p." This subproblem, of course, leads nowhere, since p is not a universally true theorem, hence cannot be proved.

At a later stage in its search LT tries the chaining method. Chaining forward, it finds the theorem "p implies (p or p)" and is then faced with the new problem of proving that "(p or p) implies p." This it is able to do by the substitution method, when it discovers the corresponding axiom.

All of these steps, successful and unsuccessful, in its proof—and the ones we have omitted from our exposition, as well—can be printed out to provide us with a complete record of how LT executed its program in solving this particular problem.

SOME CHARACTERISTICS OF THE PROBLEM-SOLVING PROCESS

Using as our data the information provided by LT as to the methods it tries, the sequence of these methods, and the theorems employed, we can ask whether its procedure shows any resemblance to the human problem-solving process as it has been described in psychological literature. We find that there are, indeed, many such resemblances, which we summarize under the following headings: set, insight, concept formation, and structure of the problem-subproblem hierachy.

Set

The term "set," sometimes defined as "a readiness to make a specified response to a specified stimulus" (4, p. 65), covers a variety of psychological phenomena. We should not be surprised to find that more than one aspect of LT's behavior exhibits "set," nor that these several evidences of set correspond to quite different underlying processes.

1. Suppose that after the program has been loaded in LT, the axioms and a sequence of problem expressions are placed in its memory. Before LT undertakes to prove the first problem expression, it goes through the list of axioms and computes a description of each for subsequent use in the "similarity" tests. For this reason, the proof of the first theorem takes an extra interval of time amounting, in fact, to about twenty seconds. Functionally and phenomenologically, this computation process and interval represent a *preparatory set* in the sense in which that term is used in reaction-time experiments. It turns out in LT that this preparatory set saves about one third of the computing time that would otherwise be required in later stages of the program.

2. *Directional set* is also evident in LT's behavior. When it is attempting a particular subproblem, LT tries first to solve it by the substitution method. If this proves fruitless, and only then, it tries the detachment method, then chaining forward, then chaining backward. Now when it searches for theorems suitable for the substitution method, it will not notice theorems that might later be suitable for detachment (different similarity tests being applied in the two cases). It attends singlemindedly to possible candidates for substitution until the theorem list has been exhausted; then it turns to the detachment method.

3. Hints and the change in behavior they induce have been mentioned earlier. Variants of LT exist in which the order of methods attempted by LT, and the choice of units in describing expressions, depend upon appropriate hints from the experimenter.

4. Effects from directional set occur in certain learning situations—as illustrated, for example, by the classical experiments of Luchins. Although LT at the present time has only a few learning mechanisms, these will produce strong effects of directional set if problems are presented to LT in appropriate sequences. For example, it required about 45 minutes to prove Theorem 2.48 in the first experiment because LT, provided with all the prior theorems, explored so many blind alleys. Given only the axioms and Theorem 2.16, LT proved Theorem 2.48 in about 15 minutes because it now considered a quite different set of possibilities.

The instances of set observable in the present program of LT are natural and unintended by-products of a program constructed to solve problems in an efficient way. In fact, it is difficult to see how we could have avoided such effects. In its simplest aspect, the problem-solving process is a search for a solution in a very large space of possible solutions. The possible solutions must be examined in *some* particular sequence, and if they are, then certain possible solutions will be examined before others. The par-

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated

ticular rule that induces the order of search induces thereby a definite set in the ordinary psychological meaning of that term.

Preparatory set also arises from the need for processing efficiency. If certain information is needed each time a possible solution or group of solutions is to be examined, it may be useful to compute this information, once and for all, at the beginning of the problem-solving process, and to store it instead of recomputing it each time.

The examples cited show that set can arise in almost every aspect of the problem-solving process. It can govern the sequence in which alternatives are examined (the "method" set), it can select the concepts that are used in classifying perceptions (the "viewing" set), and it can consist in preparatory processes (the description of axioms).

None of the examples of set in LT relate to the way in which information is stored in memory. However, one would certainly expect such set to exist, and certain psychological phenomena bear this out—the set in association experiments, and so-called "incubation" processes. LT as it now stands is inadequate in this respect.

Insight

In the psychological literature, "insight" has two principal connotations: (*a*) "suddenness" of discovery, and (*b*) grasp of the "structure" of the problem, as evidenced by absence of trial and error. It has often been pointed out that there is no necessary connection between the absence of overt trial-and-error behavior and grasp of the problem structure, for trial and error may be perceptual or ideational, and no obvious cues may be present in behavior to show that it is going on.

In LT an observer's assessment of how much trial and error there is will depend on how much of the record of its problem-solving processes the computer prints out. Moreover, the amount of trial and error going on "inside" varies within very wide limits, depending on small changes in the program.

The performance of LT throws some light on the classical debate between proponents of trial-and-error learning and proponents of "insight," and shows that this controversy, as it is usually phrased, rests on ambiguity and confusion. LT searches for solutions to the problems that are presented it. This search must be carried out in some sequence, and LT's success in actually finding solutions for rather difficult problems rests on the fact that the sequences it uses are not chosen casually but do, in fact, depend on problem "structure."

To keep matters simple, let us consider just one of the methods LT uses—proof by substitution. The number of valid proofs (of *some* theorem) that the machine can construct by substitution of new expressions for the variables in the axioms is limited only by its patience in generating expressions. Suppose now that LT is presented with a problem expression to be proved by substitution. The crudest trial-and-error procedure we can imagine is for the machine to generate substitutions in a predetermined sequence that is independent of the expression to be proved, and to compare each of the resulting expressions with the problem expression, stopping when a pair are identical (cf. 7).

Suppose, now, that the generator of substitutions is constructed so that it is *not* independent of the problem expression—so that it tries substitutions in different sequences depending on the nature of the latter. Then, if the dependence is an appropriate one, the amount of search required on the average can be reduced. A simple strategy of this sort would be to try in the axioms only substitutions involving variables that actually appear in the problem expression.

The actual generator employed by LT is more efficient (and hence more "insightful" by the usual criteria) than this. In fact, it works backward from the problem expression, and takes into account necessary conditions that a substitution must satisfy if it is to work. For example, suppose we are substituting in the axiom "p implies (q or p)," and are seeking to prove "r implies (r or r)." Working backward, it is clear that *if* the latter expression can be obtained from the former by substitution at all, then the variable that must be substituted for p is r. This can be seen by examining the first variable in each expression, without considering the rest of the expression at all (cf. 7).

Trial and error is reduced to still smaller proportions by the method for searching the list of theorems. Only those theorems are extracted from the list for attempted substitution which are "similar" in a defined sense to the problem expression. This means, in practice, that substitution is attempted in only about ten per cent of the theorems. Thus a trial-and-error search of the theorem list to find theorems similar to the problem expression is substituted for a trial-and-error series of attempted substitutions in each of the theorems.

In these examples, the concept of proceeding in a "meaningful" fashion is entirely clear and explicit. Trial-and-error attempts take place in some "space" of possible solutions. To approach a problem "meaningfully" is to have a strategy that either permits the search to be limited to a smaller subspace, or generates elements of the space in an order that makes probable the discovery of one of the solutions early in the process.

We have already listed some of the most important elements in the program of LT for reducing search to tolerable proportions. These are: (*a*) the description programs to select theorems that are "likely" candidates for substitution attempts; (*b*) the process of working backwards, which uses information about the goal to rule out large numbers of attempts without actually trying them. In addition to these, the executive routine may select the sequence of subproblems to be worked on in an order that takes up "simple" subproblems first.

Concepts

Most of the psychological research on concepts has focused on the processes of their formation. The current version of LT is mainly a performance program, and hence shows no concept formation. There is in the program, however, a clearcut example of the use of concepts in problem solving. This is the routine for describing theorems and searching for theorems "similar" to the problem expression or some part of it in order to attempt substitutions, detachments, or chainings. All theorems having the same description exemplify a common concept. We have, for example, the concept of an expression that has a single variable, one argument place on its left side, and two argument places on its right side: "p implies (p or p)" is an expression exemplifying this concept; so is "q implies (q implies q)."

The basis for these concepts is purely pragmatic. Two expressions having the same description "look alike" in some undefined sense; hence, if we are seeking to prove one of them as a theorem, while the other is an axiom or theorem already proved, the latter is likely construction material for the proof of the former.

Hierarchies of Processes

Another characteristic of the behavior of LT that resembles human problem-solving behavior is the hierarchical structure of its processes. Two kinds of hierarchies exist, and these will be described in the next two paragraphs.

In solving a problem, LT breaks it down into component problems. First of all, it makes three successive attempts: a proof by substitution, a proof by detachment, or a proof by chaining. In attempting to prove a theorem by any of these methods, it divides its task into two parts: first, finding likely raw materials in the form of axioms or theorems previously proved; second, using these materials in matching. To find theorems similar to the problem expression, the first step is to compute a description of the problem expression; the second step is to search the list of theorems for expressions with the same description. The description-computing program divides, in turn, into a program for computing the number of levels in the expression, a program for computing the number of distinct variables, and a program for computing the number of argument places.

LT has a second kind of hierarchy in the generation of new expressions to be proved. Both the detachment and chaining methods do not give proofs directly but, instead, provide new alternative expressions to prove. LT keeps a list of these subproblems, and, since they are of the same type as the original problem, it can apply all its problem-solving methods to them. These methods, of course, yield yet other subproblems, and in this way a large network of problems is developed during the course of proving a given logic expression. The importance of this type of hierarchy is that it is not fixed in advance, but grows in response to the problem-solving process itself, and shows some of the flexibility and transferability that seem to characterize human higher mental processes.

The problem-subproblem hierarchy in LT's program is quite comparable with the hierarchies that have been discovered by students of human problem-solving processes, and particularly by de Groot in his detailed studies of the thought methods of chess players (2, pp. 78–83, 105–111). Our earlier discussion of insight shows how the program structure permits an efficient combination of trial-and-error search with systematic use of experience and cues in the total problem-solving process.

SUMMARY OF THE EVIDENCE

We have now reviewed the principal evidence that LT solves problems in a manner closely resembling that exhibited by humans in dealing with the same problems. First, and perhaps most important, it is in fact capable of finding proofs for theorems—hence incorporates a system of processes that is sufficient for a problem-solving mechanism. Second, its ability to solve a particular problem depends on the sequence in which problems are presented to it in much the same way that a human subject's behavior depends on this sequence. Third, its behavior exhibits both preparatory and directional set. Fourth, it exhibits insight both in the sense of vicarious trial and error leading to "sudden" problem solution, and in the sense of employing heuristics to keep the total amount of trial and error within reasonable bounds. Fifth, it employs simple concepts to classify the expressions with which it deals. Sixth, its program exhibits a complex organized hierarchy of problems and subproblems.

COMPARISON WITH OTHER THEORIES

We have proposed a theory of the higher mental processes, and have shown how LT, which is a particular exemplar of the theory, provides an explanation for the processes used by humans to solve problems in symbolic logic. What is the relation of this explanation to others that have been advanced?

Associationism

The broad class of theories usually labelled "associationist" share a generally behaviorist viewpoint and a commitment to reducing mental functions to elementary, mechanistic neural events. We agree with the associationists that the higher mental processes can be performed by mechanisms—indeed, we have exhibited a specific set of mechanisms capable of performing some of them.

We have avoided, however, specifying these mechanisms in neurological or pseudo-neurological terms. Problem solving—at the information-processing level at which we have described it—has nothing specifically "neural" about it, but can be performed by a wide class of mechanisms, including both human brains and digital computers. We do not believe that this functional equivalence between brains and computers implies any structural equivalence at a more minute anatomical level (e.g., equivalence of neurons with circuits). Discovering what neural mechanisms realize these information-processing functions in the human brain is a task for another level of theory construction. Our theory is a theory of the information processes involved in problem solving, and not a theory of neural or electronic mechanisms for information processing.

The picture of the central nervous system to which our theory leads is a picture of a more complex and active system than that contemplated by most associationists. The notions of "trace," "fixation," "excitation," and "inhibition" suggest a relatively passive electrochemical system (or, alternatively, a passive "switchboard"), acted upon by stimuli, altered by that action, and subsequently behaving in a modified manner when later stimuli impinge on it.

In contrast, we postulate an information-processing system with large storage capacity that holds, among other things, complex strategies (programs) that may be evoked by stimuli. The stimulus determines what strategy or strategies will be evoked; the content of these strategies is already largely determined by the previous experience of the system. The ability of the system to respond in complex and highly selective ways to relatively simple stimuli is a consequence of this storage of programs and this "active" response to stimuli. The phenomena of set and insight that we have already described and the hierarchical structure of the response system are all consequences of this "active" organization of the central processes.

The historical preference of behaviorists for a theory of the brain that pictured it as a passive photographic plate or switchboard, rather than as an active computer, is no doubt connected with the struggle against vitalism. The invention of the digital computer has acquainted the world with a device—obviously a mechanism—whose response to stimuli is clearly more complex and "active" than the response of more traditional switching networks. It has provided us with operational and unobjectionable interpretations of terms like "purpose," "set," and "insight." The real importance of the digital computer for the theory of higher mental processes lies not merely in allowing us to realize such processes "in the metal" and outside the brain, but in providing us with a much profounder idea than we have hitherto had of the characteristics a mechanism must possess if it is to carry out complex information-processing tasks.

Gestalt Theories

The theory we have presented resembles the associationist theories largely in its acceptance of the premise of mechanism, and in few other respects. It resembles much more closely some of the Gestalt theories of problem solving, and perhaps most closely the theories of "directed thinking" of Selz and de Groot. A brief overview of Selz's conceptions of problem solving, as expounded by de Groot, will make its relation to our theory clear.

1. Selz and his followers describe problem solving in terms of processes or "operations" (2, p. 42). These are clearly the counterparts of the basic processes in terms of which LT is specified.

2. These operations are organized in a strategy, in which the outcome of each step determines the next (2, p. 44). The strategy is the counterpart of the program of LT.

3. A problem takes the form of a "schematic anticipation." That is, it is posed in some such form as: Find an X that stands in the specified relation R to the given element E (2, pp. 44–46). The counterpart of this in LT is the problem: Find a *sequence of sentences* (X) that stands in the relation of *proof* (R) to the given *problem expression* (E). Similarly, the subproblems posed by LT can be described in terms of schematic anticipations: for example, "Find an expression that is 'similar' to the expression to be proved." Many other examples can be supplied of "schematic anticipations" in LT.

4. The method that is applied toward solving the problem is fully specified by the schematic anticipation. The counterpart in LT is that, upon receipt of the problem, the executive program for solving logic problems specifies the next processing step. Similarly, when a subproblem is posed—like "prove the theorem by substitution"—the response to this subproblem is the initiation of a corresponding program (here, the method of substitution).

5. Problem solving is said to involve (*a*) finding means of solution, and (*b*) applying them (2, pp. 47–53). A counterpart in LT is the division between the similarity routines, which find "likely" materials for a proof, and the matching routines, which try to use these materials. In applying means, there are needed both *ordering* processes (to assign priorities when more than one method is available) and *control* processes (to evaluate the application) (2, p. 50).

6. Long sequences of solution methods are coupled together. This coupling may be *cumulative* (the following step builds on the result of the preceding) or *subsidiary* (the previous step was unsuccessful, and a new attempt is now made) (2, p. 51). In LT the former is illustrated by a successful similarity comparison followed by an attempt at matching; the latter by the failure of the method of substitution, which is then followed by an attempt at detachment.

7. In cumulative coupling, we can distinguish *complementary* methods from *subordinated* methods (2, p. 52). The former are illustrated by successive substitutions and replacements in successive elements of a pair of logic expressions. The latter are illustrated by the role of matching as a subordinate process in the detachment method.

We could continue this list a good deal further. Our purpose is not to suggest that the theory of LT can or should be translated into the language of "directed thinking." On the contrary, the specification of the program for LT clarifies to a considerable extent notions whose meanings are only vague in the earlier literature. What the list illustrates is that the processes that we observe in LT are basically the same as the processes that have been observed in human problem solving in other contexts.

PERFORMANCE AND LEARNING

LT is primarily a performance machine. That is to say, it solves problems rather than learning how to solve problems. However, although LT does not learn in all the ways that a human problem solver learns, there are a number of important learning processes in the program of LT. These serve to illustrate some, but not all, of the forms of human learning.

Learning in LT

By learning, we mean any more or less lasting change in the response of the system to successive presentations of the same stimulus. By this definition—which is the customary one—LT does learn.

1. When LT has proved a theorem, it stores this theorem in its memory. Henceforth, the theorem is available as material for the proof of subsequent theorems. Therefore, whether LT is able to prove a particular theorem depends, in general, on what theorems it has previously been asked to prove.

2. LT remembers, during the course of its attempt to prove a theorem, what subproblems it has already tried to solve. If the same subproblem is obtained twice in the course of the attempt at a proof, LT will remember and will not try to solve it a second time if it has failed a first.

3. In one variant, LT remembers what theorems have proved useful in the past in conjunction with particular methods and tries these theorems first when applying the method in question. Hence, although its total repertory of methods remains constant, it learns to apply particular methods in particular ways.

These are types of learning that would certainly be found also in human problem solvers. There are other kinds of human learning that are not yet represented in LT. We have already mentioned one—acquiring new methods for attacking problems. Another is modifying the descriptions used in searches for similar theorems, to increase the efficiency of those searches. The latter learning process may also be regarded as a process for concept formation. We have under way a number of activities directed toward incorporating new forms of learning into LT, but we will postpone a more detailed discussion of these until we can report concrete results.

What is Learned

The several kinds of learning now found in LT begin to cast light on the pedagogical problems of "what is learned?" including the problems of transfer of training. For example, if LT simply stored proofs of theorems as it found these, it would be able to prove a theorem a second time very rapidly, but its learning would not transfer at all to new theorems. The storage of *theorems* has much broader transfer value than the storage of *proofs*, since, as already noted, the proved theorems may be used as stepping stones to the proofs of new theorems. There is no mystery here in the fact that the transferability of what is learned is dependent in a very sensitive way upon the form in which it is learned and remembered. We hope to draw out the implications, psychological and pedagogical, of this finding in our subsequent research on learning.

CONCLUSION

We should like, in conclusion, only to draw attention to the broader implications of this approach to the study of information-processing systems. The heart of the approach is describing the behavior of a system by a well specified program, defined in terms of elementary information processes. In this approach, a specific program plays the role that is played in classical systems of applied mathematics by a specific system of differential equations.

Once the program has been specified, we proceed exactly as we do with traditional mathematical systems. We attempt to deduce general properties of the system from the program (the equations); we compare the behavior predicted from the program (from the equations) with actual behavior observed in experimental or field settings; we modify the program (the equations) when modification is required to fit the facts.

The promise of this approach is several-fold. First, the digital computer provides us with a device capable of realizing programs, and hence, of actually determining what behavior is implied by a program under various environmental conditions. Second, a program is a very concrete specification of the processes, and permits us to see whether the processes we postulate are realizable, and whether they are sufficient to produce the phenomena. The vaguenesses that have plagued the theory of higher mental processes and other parts of psychology disappear when the phenomena are described as programs.

In the present paper we have illustrated this approach by beginning the construction of a thoroughly operational theory of human problem solving. There is every reason to believe that it will prove equally fruitful in application to the theories of learning, of perception, and of concept formation.

REFERENCES

1. BRUNER, J. S., GOODNOW, J., & AUSTIN, G. *A study of thinking*. New York: Wiley, 1956.
2. DE GROOT, A. *Het Denken van den Schaker*. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij, 1946.
3. HILBERT, D., & ACKERMANN, W. *Principles of mathematical logic*. New York: Chelsea, 1950.
4. JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harper, 1955.
5. MOORE, O. K., & ANDERSON, S. B. Search behavior in individual and group problem solving. *Amer. sociol. Rev.*, 1955, **19**, 702–714.
6. NEWELL, A., & SHAW, J. C. Programming the logic theory machine. *Proceedings Western Joint Computer Conference* (Institute of Radio Engineers), 1957, 230–240.
7. NEWELL, A., SHAW, J. C., & SIMON, H. A. Empirical explorations with the logic theory machine. *Proceedings Western Joint Computer Conference* (Institute of Radio Engineers), 1957, 218–230.
8. NEWELL, A., & SIMON, H. A. The logic theory machine: A complex information processing system. *Transactions on information theory* (Institute of Radio Engineers), 1956, Vol. IT-2, No. 3, 61–79.
9. SIMON, H. A., & NEWELL, A. Models, their uses and limitations. In L. D. White (Ed.), *The state of the social sciences*. Chicago: Univer. Chicago Press, 1956. Pp. 66–83.
10. WHITEHEAD, A. N., & RUSSELL, B. *Principia mathematica*. Vol. I. (2nd ed.) Cambridge: Cambridge Univer. Press, 1925.

(Received July 10, 1957)

Psychological Review
Vol. 65, No. 3, 1958

AVOIDANCE CONDITIONING THROUGH IRRADIATION: A NOTE ON PHYSIOLOGICAL MECHANISMS AND PSYCHOLOGICAL IMPLICATIONS

JACK ARBIT[1]

U. S. Army Leadership Human Research Unit, Fort Ord, California[2]

Recent experiments by Garcia et al. (11, 13, 14) appear to be of great interest to psychologists studying that aspect of learning called avoidance behavior. These studies have shown that exposure to Co^{60} acts as an unconditioned stimulus (US) to produce avoidance behavior to the cues associated with the presentation of the radiation. In a footnote to one of these reports it was noted that "the physiological mechanisms through which radiation is capable of acting as an unconditioned stimulus are as yet unknown" (13, p. 40). It is the purpose of this note to indicate possible physiological mechanisms through which this effect occurs and to comment upon the importance of this phenomenon for psychology.

Wynne and Solomon (22) have shown that interfering with the normal functioning of the peripheral autonomic nervous system (ANS) by surgical and pharmacological techniques interferes with an animal's ability to acquire an avoidance response which, when learned, in some cases undergoes spontaneous extinction. Auld (3) reports that tetraethylammonium, a drug which blocks impulses through the autonomic ganglia, has a depressive effect upon speed of response in an escape-avoidance situation, although the possibility of a confounding effect due to a depressant action of the drug upon motor activity has been noted by Brady (4). By using a serial discrimination problem and utilizing such measures as errors, trials to a criterion, and trials to extinction, one may evaluate the effects of autonomic block that are independent of the effects of the drug upon activity or motor capacity of the organism. Under these conditions it was found that tetraethylammonium as well as hexamethonium, a second drug which has as one of its major effects the block of impulses through the autonomic ganglia, interferes not only with learning, utilizing escape from shock as motivation (2), but also interferes to some extent with learning in less stressful situations (1). From these studies one may conclude, although tentatively because of a shortage of confirmatory evidence, that the cues associated with the normal functioning of the ANS are important in goal directed, integrated behavior. Whether this is true because of an increased drive state produced by ANS functioning (for a recent summary see 21) or because of the feedback to the cortex from the ANS (5) is of little import for the considerations presented in this paper.

In a thorough review of the relationships between irradiation injury and host defense mechanisms, Donaldson and Marcus (7) have shown that irradiation-induced stress produces an increase in adrenal cortical function. This conclusion is based upon the work of Dougherty and White (8), Patt et al. (20), North and Nims (19), and Hochman and Bloch-Frankenthal (15), who found decreases in ascorbic acid after irradiation. Nims and Sutton (18) present data which indicate that the decrease in adrenal cholesterol level as well as the polydipsia which accompanies whole body radiation is the result of increased activity of the pituitary-adrenal system. Lasser and Stenstrom (16), working with humans following pelvic area irradiation, concluded that the adrenal cortex underwent definite changes in the course of irradiation, although these changes did not appear to be correlated with clinical "radiation sickness."[3]

---

[1] The author would like to thank Ernest Furchtgott and John Garcia for their comments on an early draft of this note.

[2] This paper was written while the author was on active duty with the U. S. Army. The opinions and conclusions contained in this report are those of the author and should in no way be construed as reflecting the views or endorsement of the Department of the Army. Present address: V. A. Research Hospital, Chicago.

If, then, the feedback associated with ANS functioning induced by irradiation has the same properties as the autonomic feedback associated with electric shock, a sudden loud noise, or any number of other stimuli, we would hypothesize that, just as the pairing of environmental cues with electric shock causes these cues to acquire fear-inducing properties, so the pairing of environmental cues with irradiation should have this effect. It is these cues, then, which, when later presented without electric shock or irradiation, produce avoidance behavior.

Gastric changes, general malaise, and nausea are also known to accompany irradiation (Furchtgott;[4] Davis et al., 6; Leary, 17) and may well cause ANS feedback as well as feedback through the proprioceptive fibers of the somatic nervous system. This feedback also may become associated with environmental cues in such a way that upon presentation of the environmental cues an avoidance response to these cues is manifest.[5]

The impact, from the psychologist's point of view, of the studies by Garcia et al. is twofold. First, it appears that a US may produce avoidance conditioning even when the subject is not aware directly of its action. Awareness in this sense refers to the functioning of particular peripheral somatic afferents. When using electric shock to the feet of a rat through a grid upon which it stands, pain receptors in the feet supply awareness of the location of the trauma. When the US is radiation, particularly at minimal dosages, the animal has no awareness of the trauma through the somatic nervous system. Some knowledge, however, that it is undergoing a stress may be mediated through the action of the ANS, the changes in gastric functioning, and the general malaise and nausea produced by irradiation.

Second, the use of radiation as a US provides a solution to the problem noted by Solomon and Brush (21) of the general lack in experimental psychology of aversive stimuli other than electric shock. In this area the implications of the studies by Garcia et al. for further research are vast indeed. In a T maze will rats learn to turn away from a goal box in which they have been irradiated? What is the function of the curve relating drive strength and dosage level? What are the temporal parameters associated with the use of radiation as a US? Will a rat learn an instrumental response, such as bar pressing, to turn off a radiation source?[6] A review of the behavioral effects of ionizing radiation may be found in Furchtgott (9). The relative importance of radiation as an aversive stimulus as compared to, let us say, bright light or intense sound, appears to be in the greater potentialities it offers for elucidating the role of the underlying physiological mechanisms in learning and the problem of learning without "awareness."

---

[3] Due to differences in dosage and exposure time, these studies are not directly applicable to the work of Garcia et al. However, it will be assumed tentatively that the time-dose parameter used by Garcia et al. results in ANS effects similar to those produced in those studies concerned more explicitly with irradiation and ANS changes.

[4] Personal communication from E. Furchtgott, University of Tennessee, May, 1957.

[5] In a recent study, Garcia and Kimeldorf (10) also reach the conclusion that the physiological mechanism underlying avoidance conditioning through irradiation may be stimulation of the ANS, although their evidence is somewhat different from that presented in this paper.

[6] Since writing this paper, it has come to the author's attention that a program of research at the U. S. Naval Radiological Defense Laboratory has been undertaken with just such questions in mind (10, 12).

REFERENCES

1. ARBIT, J. Effects of drive strength and chemical block of autonomic impulses upon maze learning and retention. *Psychol. Rep.*, 1957, **3**, 91–94.
2. ARBIT, J. Shock motivated serial discrimination learning and the chemical block of autonomic impulses. *J. comp. physiol. Psychol.*, 1958, **51**, 199–201.
3. AULD, F., JR. The effects of tetraethylammonium on a habit motivated by fear. *J. comp. physiol. Psychol.*, 1951, **44**, 565–574.
4. BRADY, J. V. Does tetraethylammonium reduce fear? *J. comp. physiol. Psychol.*, 1953, **46**, 307–310.
5. DARROW, C. W. Psychological and psychophysiological significance of the electroencephalogram. *Psychol. Rev.*, 1947, **54**, 156–168.
6. DAVIS, R. T., MCDOWELL, A. A., DETER, C. W., & STEELE, J. P. Performance of rhesus monkeys on selected laboratory tasks presented before and after a large single dose of whole-body X radiation. *J. comp. physiol. Psychol.*, 1956, **49**, 20–26.
7. DONALDSON, D. M., & MARCUS, S. Aspects of the relationship between irradiation injury and mammalian host defense mechanism: A review. *Air Univer., Sch. Aviation Med.*, 1956, No. 56–50.
8. DOUGHERTY, T. F., & WHITE, A. Pituitary-adrenal cortical control of lymphocyte structure and function as revealed by experimental X-radiation. *Endocrinology*, 1946, **39**, 370–385.
9. FURCHTGOTT, E. Behavioral effects of ionizing radiations. *Psychol. Bull.*, 1956, **53**, 321–334.
10. GARCIA, J., & KIMELDORF, D. J. Temporal relationship within the conditioning of a saccharin aversion through radiation exposure. *J. comp. physiol. Psychol.*, 1957, **50**, 180–183.
11. GARCIA, J., KIMELDORF, D. J., & HUNT, E. L. Conditioned responses to manipulative procedures resulting from exposure to gamma radiation. *Radiation Res.*, 1956, **5**, 79–87.
12. GARCIA, J., KIMELDORF, D. J., & HUNT, E. L. Spatial avoidance behavior in the rat as a result of exposure to ionizing radiation. *Brit. J. Radiol.*, 1957, **30**, 318–321.
13. GARCIA, J., KIMELDORF, D. J., HUNT, E. L., & DAVIES, B. P. Food and water consumption of rats during exposure to gamma radiation. *Radiation Res.*, 1956, **4**, 33–41.
14. GARCIA, J., KIMELDORF, D. J., & KOELLING, R. A. Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, 1955, **122**, 157–158.
15. HOCHMAN, A., & BLOCH-FRANKENTHAL, L. The effect of low and high X-ray dosage on the ascorbic acid content of the suprarenal. *Brit. J. Radiol.*, 1953, **26**, 599–600.
16. LASSER, E. C., & STENSTROM, K. W. Radiation sickness: A study of its relation to adrenal cortical function and the absolute eosinophil count. *Amer. J. Roentgenol.*, 1954, **72**, 474–487.
17. LEARY, R. W. Food-preference changes of monkeys subjected to low-level irradiation. *J. comp. physiol. Psychol.*, 1955, **48**, 343–346.
18. NIMS, L. F., & SUTTON, E. Adrenal cholesterol, liver glycogen and water consumption of fasting and X-irradiated rats. *Amer. J. Physiol.*, 1954, **177**, 51–54.
19. NORTH, N., & NIMS, L. F. Time-dose study of biochemical responses of rats to X-radiation. *Federation Proc.*, 1949, **8**, 119–120.
20. PATT, H. M., SWIFT, M. N., TYREE, E. B., & JOHN, E. S. Adrenal response to total body X-radiation. *Amer. J. Physiol.*, 1947, **150**, 480–487.
21. SOLOMON, R. L., & BRUSH, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1956.
22. WYNNE, L. C., & SOLOMON, R. L. Traumatic avoidance learning: Acquisition and extinction in dogs deprived of normal peripheral autonomic function. *Genetic Psychol. Monogr.*, 1955, **52**, 241–284.

(Received May 31, 1957)

Psychological Review
Vol. 65, No. 3, 1958

THE PROBLEM OF REDUCTIONISM IN PSYCHOLOGY

RICHARD JESSOR[1]
University of Colorado

The recent upsurge of interest in physiological determinants and physical models of behavior has raised or reopened certain fundamental questions about the logical status of psychology as an autonomous discipline among the sciences. Some of the discussions may fairly be summarized as implying that physiological or physical concepts are in some sense more basic than those of psychology and that, therefore, causal explanation of behavior will utimately be expressed in those terms. The controversy over the nature of the hypothetical construct (HC) in psychological explanation is a current example. Krech (16) has insisted that HCs must have neurophysiological reference or locus, and that psychologists should be content to define such things as needs, tensions, or cognitive structures as neural events. While this particular view has been criticized (2, 15), and there is much agreement that neurophysiological content is irrelevant in HCs (20), the more general orientation—that fundamental explanation will be reductive—continues to be influential.

Psychologists are not alone in being concerned about this problem. Anthropologists such as Kroeber and White have felt called upon to defend the autonomy of their discipline against those who have tried to explain culture in terms of the concepts of psychology or even biology. (White caustically comments that some anthropologists "have sold their culturological birthright for a mess of psychiatric pottage" [25, p. xix]). In turn, prominent biologists such as Needham, Haldane, Woodger, and Bertalanffy have argued that their own discipline cannot be considered simply an application of chemistry and physics.

Obviously, an inherent attraction of reductive explanation is its implications for possible ways of unifying the separate scientific disciplines. Since the unity of science is an ultimate aim of many scientific workers, the reductive point of view is not likely to be abandoned in the absence of reasonable alternative approaches to that goal. A commitment to reductionism, either because of its implications for the unity of science, or because of the belief that it represents more fundamental explanation, undoubtedly influences the strategy of work of many scientists.[2] For this reason alone it would seem worth while to assess the doctrine in some detail.

Some considerations of the reductionism problem have tended to dismiss it. One basis for dismissal is the assertion that the answer to the problem is entirely empirical in nature, depending on the course of future developments of science about which speculation is admittedly dangerous. This view seems unwarranted; while the ultimate relations among the sciences will be an empirical outcome, at any given point in time, the relations among the sciences are legitimate and important questions for logical analysis. Another basis for dismissal, especially where the disciplines involved are psychology and physiology, is the adoption of the reductionistic view as a logical or an in-principle certainty. The soundness of this view, also, is open to question, as will be shown. The purpose of this paper, then, is to examine the problem of reductionism and to try to make a small beginning in separating issues which are logical in character from those which are empirical. In the course of our analysis, it will be contended that there are certain *logical* barriers to any present-day physiological reductionism.

---

[1] The author was a Social Science Research Council postdoctoral fellow at the University of California during the time in which this paper was prepared.

[2] Sound empirical work, of course, requires no defense, whether motivated by reductionistic aims or not. Nothing said in this paper should be interpreted as depreciating the value of empirical or theoretical efforts to bridge the gap between disciplines.

THE DOCTRINE OF REDUCTIONISM

The essence of reductionism would seem to include four related general propositions. (*a*) The several disciplines or sciences may be considered as hierarchically ordered from, e.g., physics at the base through chemistry, biology, and psychology, to the social and historical disciplines at the top.[3] (*b*) The second essential aspect of reductionism is the proposition that the terms or concepts and the relations or laws of one discipline may fully and without loss of meaning be translated into or deduced from those of another discipline. (*c*) Such deduction or derivability proceeds only in one direction, from lower to higher[4] levels in the hierarchical ordering, and hence the term "reductionism"; terms and laws of the higher discipline are "reduced" to those of a lower one. Thus, in our earlier example, the psychological term "cognitive structure" is considered translatable into—deducible from—terms belonging to neurophysiology. (*d*) The final aspect is the implicit or explicit proposition that the lower the level of terms employed to explain a given phenomenon, the more causal or fundamental or basic the explanation. This is really only a corollary of the first point if certain assumptions about the nature of the hierarchical ordering are made. These four propositions together would seem to constitute the essential meaning of reductionism as a general doctrine. An adherent of that point of view may not, of course, subscribe to all of its aspects.

The primary focus of this paper is upon the issues attending the reduction of psychology to physiology, and our evaluation of the doctrine will for the most part be oriented toward that specific context. Within that context, the following comments are illustrative of the position which supports the doctrine. "Logically and in principle, physiological reduction is a certainty.

---

[3] Finer discriminations can, of course, easily be made by including the well-known "border" disciplines such as biochemistry or social psychology. But the fact that there is no sharp break between the sciences, and that it is frequently difficult to tell where one leaves off and the other begins, need not, in itself, challenge the autonomous existence of the several disciplines. What may be implied by ordering them in a hierarchy from lower to higher will be discussed shortly; for the moment it is only important to consider that this is one of the notions essential to the doctrine of reductionism.

[4] Terms in this paper referring to "position" in the hierarchy of the sciences, e.g., higher-lower, upward-downward, top-bottom, are by no means to imply any valuative judgment. The *meaning* of position in the hierarchy has been variously specified, for example, as referring to levels of abstraction or levels of integration, or as referring to the order of historical evolution within the universe of the subject matter of the sciences, or even to the order of historical emergence of the sciences themselves (25). As pointed out later, the concept of levels of science is not an analytically clear one (cf. 17). For present purposes it is sufficient to take note of the existence in scientific discourse of such a hierarchical concept, and to recognize the traditional general ordering which places the physical sciences at the base, the biological sciences in the middle, and the social sciences at the top of the hierarchy.

Every bit of behavior and everything that can, like conscious contents, be defined in terms of behavior has its physiological correlate" (2, p. 442); "Relative to the 'molar' (or macro-) account given by behavioristic psychology, the neurophysiological account is a micro-description of the very same events and processes" (10, p. 623); and finally, ". . . molar behavioristics is in theory completely reducible to underlying neurophysiological principles. . . . A completely deterministic neurophysiology must of necessity permit derivation of all molar behavioral laws" (20, pp. 261–262). An obvious corollary of these statements is the logical reducibility of physiological principles in turn to those of chemistry and ultimately of physics.

Any challenge to these kinds of statements and their implications can be seen to require both logical and empirical arguments. We shall try to show where each kind is appropriate.

To begin with, to speak of reducing one discipline to another requires that the terms or concepts of the one be distinguishable from those of the other. This is not an easy requirement. Woodger (26) calls it an interesting methodological question to inquire how one knows what belongs to the language of neurology and what belongs to that of psychology. Similarly, in commenting on the new terms which will accrue as science develops, Hempel (12) notes that it is by no means certain that each of these terms will be readily classifiable as physical or nonphysical. What seems to be a necessary preliminary for our examination of physiological reductionism is some general criterion for separating or identifying the terms of psychology and physiology. If this is accomplished by defining psychology in a certain way, it should make apparent some logical barriers to the possibility of a physiological reduction.

A FUNCTIONAL DEFINITION OF PSYCHOLOGY

Admittedly, the definition of any scientific discipline is somewhat arbitrary; despite this, it is certainly possible to obtain adequate agreement on criteria for segregating one discipline from another or for grossly circumscribing the domain of a particular science. Among psychologists there is considerable agreement that the scope and subject matter of concern is the behavior of whole, human[5] organisms. The difficulty with such a general statement is that the term "behavior" is not without ambiguity; psychologists have been notoriously neglectful in providing a systematic definition of a response. This laxness has, it is felt, obscured the conceptual boundary between psychology and physiology. The present discussion of behavior as a psychological concept follows the implicit orientation of all functional behavior theories, and more specifically, the approaches of Kantor (14), Brunswik (5, 6, 7), and Tolman (24).

The central point of these approaches is that behavior, *qua* psychological, refers to an organism-environment *interaction* or *relationship*. Tolman (24) specifically states that the complete identification of any behavior-act requires reference to its relation to particular goal-objects and the intervening means-objects with which it has commerce. Kantor's interbehaviorism makes essentially the same point. In considering the question of where to establish the boundaries of a behavior, Bentley (1) similarly concludes that behavior must be recognized as a transdermal process or event whose description must immediately and functionally include the environmental and situational settings.

---

[5] The writer assumes that the study of animal behavior by psychologists is merely propaedeutic to a science of human behavior.

Within this framework, then, behavior viewed psychologically is interactional or relational in nature; its specification or identification at the referential level requires the specification of a particular context and a set of relationships thereto. Our definition of psychology, therefore, excludes the study of organisms or physical environments per se, and behavior may not be referred to either alone. The laws of behavior of a discipline so defined refer to the dynamics of organism-environment functional interaction. The terms or concepts of those laws describe what may be called an interaction or behavior space.

This definition of psychology was undertaken as a means of providing a criterion for deciding whether a behavior term, or a law involving such a term, properly belongs to psychology or physiology. Our criterion requires of psychological terms that they have immediate reference to a functionally defined environment or context. Before drawing the implications of these considerations for reductionism, a further comment on this kind of definition may be in order. Some objection may be raised to the relational or transdermal character of the definition in that there is provided no palpable locus for a psychologically defined behavior. Those who raise such a query seem to be operating within what Woodger (27) picturesquely describes as a "finger and thumb" philosophy of metaphysics, i.e., the notion that a thing is real or exists only if it can in principle be picked up between the finger and thumb. Interactions or relations, though not simple physical objects, are nevertheless real and concrete and can be precisely specified by the conditions and course of their occurrence.

INCOMPLETE DERIVABILITY OF TERMS

It may now be asserted that the reference of psychology, as defined, is profoundly different from that of physiology. The terms and laws of the latter refer to intraorganismic or intradermal processes, or, at most, relations between an organism or its parts and the space defined by physics. They refer, in short, to the functioning of anatomical structures or processes described by body parameters. To state one of the major contentions of this paper, the absence in contemporary physiology of any systematic terms for describing the functional environment or context of behavior would seem to preclude, *on logical grounds alone,* any complete reduction of psychology to physiology. These necessary and sufficient conditions for the terms of psychology cannot be described in physiological (or physical) terms alone. This "incompleteness" of the lower discipline's language, in being able to specify only the physiological correlates but not the environmental correlates of a behavior or response, constitutes the logical impediment. This point obviously requires elaboration and further support.

Let us take as an example an occurrence described in common-sense, nonsystematic language—the wave of an arm as two persons pass each other—and compare the systematic descriptions of it by the two disciplines being discussed. The systematic language of the physiologist enables him to rely on only body parameters or physical terms. He may thus speak of arm-displacement, changes in muscle-tension, metabolic rate, blood-volume distribution, and neural reactivity. In none of these terms is it systematically possible to take cognizance of the social context. On the other hand, the psychologist may describe the event as waving a greeting to a friend, or even, since micromediation is generally of little interest to him, and equifinality generally taken for granted,[6] simply as greeting a friend. If psychological laws refer to interactions between organisms and functionally defined environments, only these latter descriptions of events can logically lead to the achievement of such laws. A discipline such as physiology, lacking such contextual terms, cannot therefore be considered *logically* equivalent; hence its adequacy cannot be guaranteed, even in principle, as a complete reduction base for psychology.

One of the strongest empirical sources of support for some of these considerations may be found in the research analyses of Campbell (8, 9) and Smedslund (22, 23) evoked by the recent controversy over "what is learned." As mentioned earlier, the term "response" has been conceptually neglected; this neglect seems, in turn, to be related to the inadequate attention given to the environment or context of behavior by many psychological theorists (cf. 13 in this connection). Brunswik (5, 6, 7), however, is one of those who has called attention to the problem in his emphasis upon *distal* achievement in the adjustment of an organism to its ecology. Beginning with this orientation, and appraising a variety of learning experiments, especially those dealing with transposition, both Campbell and Smedslund conclude that the learned response must be defined in environmental terms in order successfully to accommodate—predict—the actual research findings. Campbell states that ". . . the learned response is to be essentially defined in terms of a shift in the organism-environment relationship rather than a motor response defined in terms of organism or body parameters alone" (9, p. 105).

Scrutiny of the literature on reductionism has showed only meager if any attention to the specific issue raised in this paper. Two writers may be cited whose remarks, made in other contexts, are pertinent to the logical soundness of our contention. In his Tarner lectures (26), Woodger stresses the role of environment as a determinant at various biological levels from zygote to whole human organism. He rejects the body-mind dualism in favor of speaking about persons and developing a person-language. The notion of person requires, for its very definition, environmental specification, and the latter, he observes, requires words belonging to sociology. To treat persons otherwise, i.e., in terms of body parameters, loses sight of this fact: "But this is the only way in which we can treat them so long as we confine ourselves to the physical sciences, since these sciences do not provide a vocabulary for speaking about them in any other way" (26, p. 261). Hempel has discussed the problem in relation to the possible derivability of all the laws of empirical science from those of physics (a logical corollary of physiological reductionism). Affirming that not all the terms of empirical science are definable by means of the vocabulary of physics, he asserts that ". . . a law containing, say, certain biological terms cannot, in general, be logically derived from a set of physical laws, which contain no biological terms at all . . ." (12, pp. 320–321).[7] Both of these writers seem in support of our point of view that there are logical barriers to any thorough-going physiological reductionism. The barriers reside in the absence of terms in the "lower" discipline which would enable the logical derivability of descriptions of the functional context of behavior and, thereby, the derivation of the laws of psychology.

---

[6] That is, the same psychological event may be served by (partly constituted of) an almost infinite variety of different physiological events.

[7] To achieve such a derivation requires some law connecting the biological concepts with the physical concepts. "But those connecting laws are not purely physical in character" (12, p. 321). And they have the character of *empirical* laws.

At least two kinds of questions can be raised about our analysis and are worth consideration at this point. The first of these is the possibility of overcoming the logical problem we have raised simply by incorporating into physiology the category of terms we have suggested that it lacks. This is frequently what is implied by the phrase "in principle" in assertions about the possibility of reductionism. But as Sellars (21) has pointed out in another connection, this makes the entire problem an empty truism in that it involves a "tacit redefinition" of physiological theory to encompass psychology. Such a redefinition is a statement about the *future* state of the sciences involved, and it therefore transfers the discussion from logical to empirical grounds. That is, the question of whether a future physiology will be able to encompass psychology depends entirely upon the nature and direction of ongoing empirical development of *both* disciplines. It is to be noted here, too, that the meaning of reduction is always and only relative to a given state of the disciplines concerned. To suggest adding terms to physiology in order to make psychology deducible from it implies the "elevation" or "expansion" of physiology just as much as it implies reduction of psychology. (In connection with these issues, see the Meehl-Sellars discussion of the logic of emergentism, 19.) Finally, it seems unlikely that scientific theory develops or advances by simple accretion of the terms of other theories. We will return to this point shortly, in considering reductionism and the unity of science.

The second kind of question which may be raised is implicit in the earlier quote from Bergmann. He bases his affirmation of the logical certainty of physiological reductionism on the proposition that ". . . everything that can be defined in terms of behavior has its physiological correlate" (2, p. 442). The essential point of this position would be that the environmental reference of behavior upon which we based our definition must, if effective, be represented within the organism in its physiology or, especially, neurophysiology. This is, in a sense, a proximal approach to behavior rather than a distal one. Certain arguments may be brought to bear against this position, such as the abscence of strict one-to-one proximal-distal correlation—e.g., in perception—and the theoretical significance of vicarious functioning or equifinality. The major reply which may be made, however, is that the "recovery" of actual behavioral phenomena from physiological correlates requires the conceptual coordination of these correlates to environmental contexts. Thus the issue raised originally reappears, the necessity for terms to represent or describe the context of behavior. That such coordination can or will be accomplished in the future is an empirical rather than a logical problem, and therefore not a logical certainty.

Our discussion of reductionism up to this point has concerned itself largely with the logical problems inherent in the second and third essential aspects of the doctrine as outlined at the start of this paper. Proper attention to the first and fourth aspects would extend the paper beyond practical space limitations. Instead, we shall simply sketch some of the issue requiring attention.

The first aspect has to do with the hierarchical ordering of the several scientific disciplines. Despite the widespread acceptance of the hierarchy notion—witness the frequent reference to "levels" of science, the employment of terms like "basic" to contrast disciplines, and the characterization of certain disciplines as "emergent" from others—it is not an analytically clear concept. Kroeber (17) remarks on the absence of any adequate attempt to examine systematically what the levels constitute or mean in terms of a theory of knowledge. Most of the discussions of what is meant by levels resolve into two positions which are generally considered to exclude one another. One point of view conceives the levels of science to be a matter of methodology only, i.e., to refer to the kinds of procedures employed by the various sciences, the size of their units of analysis, etc. The other point of view considers the levels to refer to substantive differences in the events or phenomena dealt with by the various disciplines. Psychologists will be familiar with this contrast from the Littman-Rosen (18) analysis of the molar-molecular problem. A third possibility is that these two positions are correlated rather than mutually exclusive, namely, that substantively different events require particularly appropriate methodological procedures for useful analysis.

Each of these three views has certain implications for the doctrine of reductionism. For example, the methodological position would seem to favor the doctrine, since it assumes the events or phenomena to be the same and only the descriptions of them to be different. If only the descriptions differ, e.g., in size of unit, it should be logically possible to reduce the larger units to their smaller constituents. On the other hand, to assert a substantive difference between levels would seem to be unfavorable to reductionism. The events or phenomena of higher levels are considered different from—not the same ones as—those of the lower levels, and therein lies the difficulty in reducing descriptions of one kind of event to those of another kind. No vitalistic or dualistic considerations need be involved in speaking of events as different; certainly organic events may in general be separated from inorganic ones, for example. Feigl's double-language theory (10) of the mind-body problem espouses the methodological position in insisting that the factual reference of the mentalistic, behavioristic and neurophysiological languages or levels of description is identical, i.e., involves the very same events and processes. Others concerned with the problem seem to adopt the substantive position; thus White (25) speaks of culture as a distinct class of events, a distinct order of phenomena. Woodger tentatively concludes that "... perhaps, in spite of superficial appearances, person-acts and behaviour [defined in physical or physiological terms] are not quite the same things ..." (26, p. 284). The position taken by the present writer in differentiating psychological from physiological terms may be seen as compatible with the substantive view. What may account for substantive differences, whether the substantive vs. methodological distinction itself is useful or defensible, the issues involved in emergentism, and any fuller analysis of the implications of the hierarchy of sciences for reductionism must be deferred for discussion elsewhere.

The final aspect of the reductionism doctrine has to do with the idea that causal explanation is advanced by the employment of terms of a lower-level discipline. In psychology, this notion identifies causal explanation with neurophysiological reference. The key reason for this approach would seem to be the belief in the higher levels as simply derivable from, or applications of, terms and laws of lower disciplines. Once this idea is abandoned, causal explanation could just as logically proceed upward to sociological and anthropological concepts. Knowledge of causality is probably best divorced from the hierarchy of sciences notion and considered instead to vary with the scope of the network in which any concept is embedded.

We have tried to examine some of the logical and empirical problems related to the doctrine of physiological reductionism. Some of the considerations have led us to doubt that psychology can be reduced to physiology, though certainly many of the questions are of an empirical sort. Nothing thus far asserted in any way denies the possibility or the desirability of the unification of the sciences, the synthesis of psychology and physiology. What is questioned is that such synthesis must proceed by reduction, by one discipline devouring or incorporating the other. All of the sciences are developing, and their influence upon each other does not only proceed upward from physics. Bertalanffy (3, 4), for example, sees certain biological developments, such as the notions of open systems, requiring extensions of the conceptual system of physics. Unity of science for him refers only to the structural isomorphy of laws in the different fields of science, the approach of General Systems Theory. Psychologists can also see the conceptual system of physiology as influenced by the data and concepts of psychology (11). Thus, unification of the sciences may proceed from above as well as from below. The continued autonomous development of each of the sciences will at least serve to specify the properties required of any synthetic unifying scheme. This is probably part of what Brunswik had in mind when he wrote: "Insistence on reduction as a universal goal of science can only result in blighted spots on the landmap of scientific enterprise" (7, p. 237).

SUMMARY

This paper has had as its aim the instigation of renewed attention to the doctrine of reductionism, especially in terms of its implications for the relationship of physiology and psychology. Despite the empirical character of the ultimate answer, it is asserted that the questions involved in the doctrine may properly be the concern of a logical analysis. After briefly sketching four propositions which constitute the essential notions of reductionism, the argument focused upon the logical possibility of a complete translatability or derivability of the concepts and laws of psychology from those of physiology. The central contention was that the latter, lacking terms to describe the behavioral environment, was logically inadequate as a base for a thoroughgoing reduction of the former. The remainder of the paper commented upon the relationship of the doctrine to the idea of a hierarchical ordering of the sciences and to the possibility of achieving a unification of science.

REFERENCES

1. BENTLEY, A. F. The behavioral superfice. *Psychol. Rev.*, 1941, **48**, 39–59.
2. BERGMANN, G. Theoretical psychology. In C. P. Stone (Ed.), *Annu. Rev. Psychol.*, 1953, **4**, 435–458.
3. BERTALANFFY, L. VON. Problems of general systems theory. *Human Biol.*, 1951, **23**, 302–312.
4. BERTALANFFY, L. VON. *Problems of life.* New York: Wiley, 1952.
5. BRUNSWIK, E. The conceptual focus of some psychological systems. *J. unif. Sci.*, 1939, **8**, 36–49.
6. BRUNSWIK, E. The conceptual framework of psychology. Chicago: Univer. Chicago Press, 1952. (*Int. Encycl. unif. Sci.*, Vol. 1, No. 10.)
7. BRUNSWIK, E. In defense of probabilistic functionalism: A reply. *Psychol. Rev.*, 1955, **62**, 236–242.
8. CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.
9. CAMPBELL, D. T. Adaptive behavior from random response. *Behavioral Sci.*, 1956, **1**, 105–110.
10. FEIGL, H. The mind-body problem in the development of logical empiricism. In H. Feigl & May Brodbeck (Eds.), *Readings in the philosophy of science.* New York: Appleton-Century-Crofts, 1953. Pp. 612–626.
11. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

12. HEMPEL, C. G. General systems theory and the unity of science. *Human Biol.*, 1951, **23**, 313–322.
13. JESSOR, R. Phenomenological personality theories and the data language of psychology. *Psychol. Rev.*, 1956, **63**, 173–180.
14. KANTOR, J. R. Preface to interbehavioral psychology. *Psychol. Rec.*, 1942, **5**, 173–193.
15. KESSEN, W., & KIMBLE, G. A. "Dynamic systems" and theory construction. *Psychol. Rev.*, 1952, **59**, 263–267.
16. KRECH, D. Dynamic systems, psychological fields, and hypothetical constructs. *Psychol. Rev.*, 1950, **57**, 283–290.
17. KROEBER, A. L. *The nature of culture.* Chicago: Univer. Chicago Press, 1952.
18. LITTMAN, R. A., & ROSEN, E. Molar and molecular. *Psychol. Rev.*, 1950, **57**, 58–65.
19. MEEHL, P. E., & SELLARS, W. The concept of emergence. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science.* Vol. I. Minneapolis: Univer. Minnesota Press, 1956. Pp. 239–252.
20. ROZEBOOM, W. W. Mediation variables in scientific theory. *Psychol. Rev.*, 1956, **63**, 249–264.
21. SELLARS, W. Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science.* Vol. I. Minneapolis: Univer. Minnesota Press, 1956. Pp. 253–329.
22. SMEDSLUND, J. The problem of "What is learned?" *Psychol. Rev.*, 1953, **60**, 157–158.
23. SMEDSLUND, J. *Multiple-probability learning.* Oslo: Akademisk Forlag, 1956.
24. TOLMAN, E. C. *Purposive behavior in animals and men.* Berkeley: Univer. California Press, 1949.
25. WHITE, L. A. *The science of culture.* New York: Farrar, Strauss, 1949.
26. WOODGER, J. H. *Biology and language.* Cambridge: Cambridge Univer. Press, 1952.
27. WOODGER, J. H. *Physics, psychology and medicine.* Cambridge: Cambridge Univer. Press, 1956.

(Received July 22, 1957)

Psychological Review
Vol. 65, No. 3, 1958

THE STATUS OF INHIBITION IN HULL'S SYSTEM: A THEORETICAL REVISION[1]

H. GWYNNE JONES

Institute of Psychiatry (Maudsley Hospital), University of London

Hull (6, 7, 8) clearly distinguishes between a habit (${}_sH_R$) and a drive state (D) and postulates that these interact in a multiplicative manner so as to generate a reaction potential (${}_sE_R$). Thus:

$${}_sE_R = f(D \times {}_sH_R)$$

In his discussion of factors affecting response decrement, Hull further postulates the development of two forms of inhibition: an innate, fatigue-produced *reactive inhibition* (I_R) and *conditioned inhibition* (${}_sI_R$).

The former is clearly conceived of and defined as a negative drive state, while the latter is considered a negative habit phenomenon. Despite this distinction, and in conflict with his earlier insistence on a multiplicative relationship between habits and drives, Hull postulates that these two variables interact in an additive fashion to produce inhibitory potential ($\dot{I}_R$). When such an inhibitory potential is operative it is subtracted from reaction potential to give a resultant or net reaction potential (${}_s\bar{E}_R$). Thus:

$${}_s\bar{E}_R = f[(D \times {}_sH_R) - (I_R + {}_sI_R)]$$

There has been considerable criticism of this formulation by such writers as Gleitman et al. (5) and Koch (3). Osgood (10) and Zeaman (15) similarly recognize the inconsistency of the equation and propose that ${}_sI_R$, as a negative habit, be subtracted directly from ${}_sH_R$, a positive habit, while continuing to allow I_R to be subtracted directly from ${}_sE_R$. Iwahara (9) recognizes that I_R, as a drive, must enter into some form of multiplicative relationship, but, in order to avoid certain implications of his argument, he then abandons ${}_sI_R$ as a habit phenomenon and proposes to consider it only as a secondary drive. Such an abandonment of the concept of a negative habit would involve a radical revision of much existing theory, while its retention is not imcompatible with the achievement of internal consistency within the Hullian system.

Iwahara bases his rejection upon only one possible modification of the basic equation, i.e., that inhibitory potential ($\dot{I}_R$) be considered as the product rather than the sum of reactive and conditioned inhibition. Symbolically:

$$\dot{I}_R = I_R \times {}_sI_R$$

He points out that this implies that the inhibitory potential would be zero when either I_R or ${}_sI_R$ is zero, a consequence clearly inconsistent with the known facts of learning. Conditioned inhibition, however, may, as a habit, also be energized by positive drives, e.g., when response decrement persists after long rest subsequent to massed practice. Similarly, reactive inhibition may interact with positive habits to lower reaction potential when no conditioned inhibition is present, as at the commencement of massed practice or extinction.

---

[1] This article is based upon a paper presented to the XV International Congress of Psychology at Brussels in August 1957. The proposed equation was earlier made use of by Eysenck (4) in a theoretical analysis of "warm-up" phenomena.

Thus it is suggested that net reaction potential may well be determined by the multiplicative interaction of a *resultant* drive state ($D - I_R$), and a *resultant* habit state ($_SH_R - {}_SI_R$), so that:

$$_S\bar{E}_R = f[(D - I_R) \times ({}_SH_R - {}_SI_R)]$$

The habit term in this equation is consistent with Osgood's and Zeaman's suggestion, while the drive term is consistent with Kimble's (**12**) experimentally validated proposition that the amount of reactive inhibition a subject will tolerate is a function of his drive level. This equation is mathematically equivalent to:

$$_S\bar{E}_R = f[(D \times {}_SH_R) - (I_R \times {}_SH_R) - (D \times {}_SI_R) + (I_R \times {}_SI_R)]$$

the fourth term of which represents a paradoxical addition to positive reaction potential from the interaction of two inhibitory variables. Similarly, if both resultant drive and habit states are negative, the resultant reaction potential would be positive. Such a state of affairs would occur rarely in practice, but may furnish a theoretical explanation of such phenomena as the "ultraparadoxical" phase of inhibition observed by Pavlov (**11**). The various inhibitory states of the organism induced by Pavlov in various ways may be considered as inhibitory drive states akin to reactive inhibition. In advanced stages of such inhibition, Pavlov noted that positive conditioned stimuli tended to lose their effect, whereas "well developed negative stimuli" (i.e., when $_SI_R$ was predominant) "acquired definite excitatory properties."

Any amendment of the Hullian equation must serve, not only to make Hullian theory internally consistent, but also to generate theorems consistent with empirical data. One such theorem, derived from the present amendment, refers to stimulus discrimination learning, around which an apparently adequate theory in Hullian terms has been developed, particularly by Spence (**8, 14**).

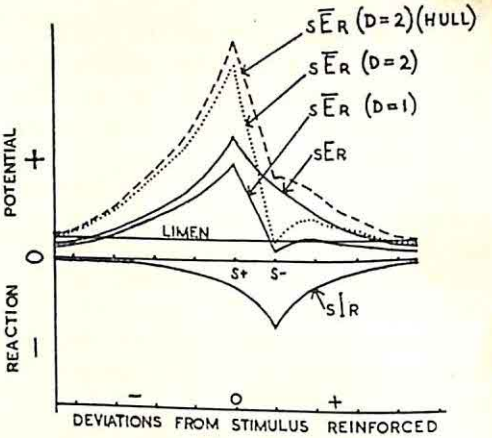


FIG. 1. Discrimination gradients at two levels of drive. See text for complete explanation.

The solid lines in Fig. 1 conventionally illustrate the algebraic summation theory of discrimination,[2] in which net reaction potential ($_S\bar{E}_R$) is a result of the subtraction of generalized inhibition ($_SI_R$) from generalized positive reaction potential ($_SE_R$). At the stage of learning illustrated, the net reaction potential attached to the nonreinforced or negative stimulus is well below the threshold for response, and the discrimination will be almost perfect.[3]

---

[2] The positive and negative generalization gradients in the figure were not calculated by formula but merely estimated, by eye, from textbook examples. Errors in reproduction and resulting differences between the two gradients will account for the unconventional hump in the curves of net reaction potentials. These are irrelevant to the argument in this paper.

[3] Hull claims that $_S\bar{E}_R$ cannot fall below $_SL_R$, the reaction threshold, because, when $_S\bar{E}_R = {}_SL_R$, no further responses occur and so no further $_SI_R$ develops. This argument neglects the effect of oscillation potential.

Let us now consider the effect of increasing drive in this situation; and, for convenience, let us assume that drive is, in fact, doubled. According to orthodox Hullian theory, this would have the effect of doubling the positive reaction potential but the conditioned inhibition would remain unchanged. The generalized net reaction potential ($_{S}\bar{E}_{R}$) would then be represented by the broken line in Fig. 1. The net potential attached to the nonreinforced stimulus is in fact higher than at the commencement of discrimination learning in the low drive situation: the discrimination would have broken down completely. Even with lesser increases of drive, its multiplicative relationship with habit strength would insure severe impairment of discrimination.

If, however, the amended equation is employed, the doubled drive would have the effect of doubling both positive and negative potentials. The generalized net potential would then be represented by the dotted line in Fig. 1 and, at any point along the stimulus dimension, would be double that in the low drive conditions. At the nonreinforced stimulus value this would involve only a slight increment in potential, as the initial value is already subthreshold. Discrimination, while somewhat impaired, would not be expected to break down completely.

Empirical evidence is clearly in favor of the prediction from the amended equation. Skinner (**13**), for example, reports that under low hunger drive rats achieve almost perfect discrimination, performing bar-pressing responses when the box is dark and not pressing when the box is lighted. Under increased food deprivation the discriminations do not break down, but become impaired in that there is some "leaking" of positive responses to the negative stimulus.

Another situation in which the alternative equations would yield different predictions as to the effects of changes in drive is that in which habit strength is equal to conditioned inhibition. Calvin et al. (**1**) carried out an experiment to show that, despite continued reinforcement and under conditions of repeated massed practice, rats will ultimately refuse to traverse a runway for food and that this inhibition is retained over lengthy periods of time without such practice. Calvin's original experiment has been severely criticized by Dinsmoor (**2**) but, in a recent unpublished study, D. C. Kendrick achieved essentially similar results, using thirst motivation and controlling the factors relevant to Dinsmoor's criticisms.

If animals, trained in this fashion, are retested under conditions of increased drive, Hull's equation indicated that a reinstatement of the running performance should occur, whereas under decreased drive the inhibition should be more effective than in the original condition. According to the amended equation, change of drive in either direction, in that it affects $_{S}H_{R}$ and $_{S}I_{R}$ equally, should have no effect on the balance between these two variables. Any change of drive, in that it alters the stimulus situation, may affect performance, but running should be no more frequent with increased than with decreased drive. In a pilot study, employing ten animals and testing this prediction, Kendrick found that three animals responded positively under increased drive. Two of these and one other responded under decreased drive. No responses occurred when

---

When $_{S}\bar{E}_{R} = {}_{S}L_{R}$, owing to oscillation, the probability of response evocation will be 0.5. Only when $_{S}\bar{E}_{R}$ falls appreciably below $_{S}L_{R}$ will this probability fall to a low level.

drive level was unchanged. A more rigorous follow-up of this finding is now in progress.

SUMMARY

An amendment of the Hullian equation is proposed which, for both excitatory and inhibitory potentials, is consistent with Hull's distinction between habit and drive states and the postulated multiplicative interaction between the two. Theorems derived from this equation are compared with the analagous consequences of Hull's formulation in the light of available empirical data.

REFERENCES

1. CALVIN, A. D., CLIFFORD, L. T., CLIFFORD, B., BOLDEN, L., & HARVEY, J. Experimental validation of conditioned inhibition. *Psychol. Rep.*, 1956, **2**, 51–56.
2. DINSMOOR, J. D. Absurdum revisited: A comment. *Psychol. Rep.*, 1956, **2**, 255–256.
3. ESTES, W. K., KOCH, S., MACCORQUODALE, K., MEEHL, P. E., MUELLER, C. G., SCHOENFELD, W. N., & VERPLANCK, W. S. *Modern learning theory.* New York: Appleton-Century-Crofts, 1954.
4. EYSENCK, H. J. "Warm-up" in pursuit rotor learning as a function of the extinction of conditioned inhibition. *Acta psychol.*, 1956, **12**, 349–370.
5. GLEITMAN, H., NACHMIAS, J., & NEISSER, U. The *S-R* reinforcement theory of extinction. *Psychol. Rev.*, 1954, **61**, 23–33.
6. HULL, C. L. *The principles of behavior.* New York: Appleton-Century-Crofts, 1943.
7. HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
8. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
9. IWAHARA, S. Hull's concept of inhibition: A revision. *Psychol. Rep.*, 1957, **3**, 9–10.
10. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
11. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
12. KIMBLE, G. A., & BILODEAU, G. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.
13. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.
14. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
15. ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received September 10, 1957)

VOL. 65, NO. 4 JULY, 1958

THE PSYCHOLOGICAL REVIEW

SYMMETRIC UNCERTAINTY ANALYSIS AND ITS IMPLICATIONS FOR PSYCHOLOGY[1]

W. R. GARNER

Johns Hopkins University

Information theory, as developed primarily by Shannon (8), has had considerable impact on psychological research. Information theory has been used as a model for many different types of behavior; it has also provided psychologists with a statistical measure which has many useful properties. Frequently the measure is used where the term *information* is somewhat inappropriate, and for this reason Garner and McGill (3) have suggested that the term *uncertainty* be applied to the measure to divorce its statistical properties from the content implications of the term *information*. Even when the mathematical properties of the uncertainty measure lead to implications for behavior situations, information theory does not necessarily define the problem; rather, the mathematical properties of the uncertainty measure simply make certain relations clear which other statistical techniques have not been able to do. The purpose of this paper is to develop some equations of this sort involved in uncertainty analysis, and to point out the relevance of these equations to a few psychological problems.

Two valuable properties of the measure of uncertainty, as pointed out by Garner and McGill (3), are: (*a*) that it is a nonparametric measure which can be applied to any set of categorized data, and (*b*) that the uncertainty of a variable can be partitioned into component parts, much as variance is partitioned in analysis of variance. A third useful property of the uncertainty measure is that an uncertainty analysis can be carried out in a completely symmetric form, in which it is unnecessary to distinguish between meanings or functions of the several variables involved in the analysis. This last property derives in part from the nonparametric nature of the measure, and is not a property of analysis of variance, in which different properties of the criterion and predictor variables are assumed.

---

[1] Parts of this paper were presented at the Fifteenth International Congress of Psychology, Brussels, 1957. This work was supported by Contract N5-ori-166, Task Order 1, between the U. S. Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-213, Project Designation No. NR 145-089, under that contract. Reproduction in whole or in part is permitted for any purpose of the United States Government. I wish to acknowledge the valuable critical comments which Alphonse Chapanis made on the manuscript.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE

ANALYSIS OF THREE VARIABLES

For illustration purposes, let us consider first a three-dimensional matrix of data. Figure 1 shows schematically such a data array and some of the terms which we will be using. The three variables are w, x, and y, and they can assume values w_i, x_j, and y_k. Each cell in the matrix originally contains a number which represents the frequency of occurrence of cases having that particular combination of values of w, x, and y. For our purposes, we have shown the matrix in which these frequencies have been transformed into proportions, $p(i,j,k)$, by dividing the cell frequencies by the total number of cases in the matrix. This matrix can be collapsed across any one variable to form three different two-variable matrices, or it can be collapsed across two variables to give distributions for one variable at a time. When two-variable matrices are used, the proportions in the cells are still computed as the frequency in a cell divided by the total number of cases, and can be obtained by summing the proportions across the collapsed variable.

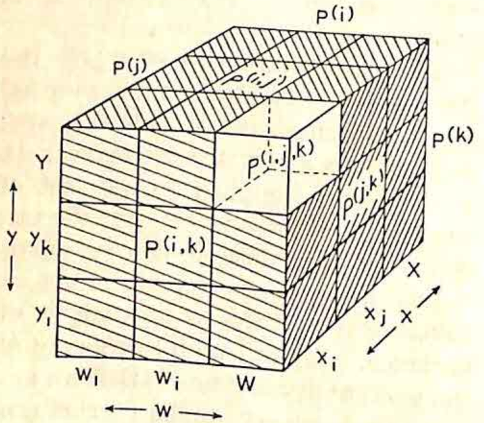


FIG. 1. Symbolic representation of a three-variable matrix of data. Three variables, w, x, y, have category values w_i, x_j, and y_k. Each cell in the matrix originally has a frequency which is divided by the total number of cases in the matrix to give the proportion $p(i,j,k)$. Summing these proportions across any variable gives two-variable matrices with cell entries of $p(i,k)$, $p(j,k)$, or $p(i,j)$ as indicated on the three surfaces. Summing across two variables gives the marginal proportions $p(i)$, $p(j)$, and $p(k)$.

Unidirectional analysis. When uncertainty is partitioned as in analysis of variance (**3**, **4**), we start by identifying one of the variables, let us say y, as the criterion variable and the other two, w and x, as predictor variables. Now our problem is to partition the uncertainty of y into its component parts. The basic partitioning equation is

$$U(y) = U(y\!:\!w,x) + U_{wx}(y) \qquad [1]$$

where $U(y)$ is the uncertainty of the y-distribution, based on $p(k)$; $U(y\!:\!w,x)$ is the total uncertainty in y which is predictable from w and x, and it is called a multiple contingent uncertainty; and $U_{wx}(y)$ is the residual or error uncertainty.[2] This last term is a conditional uncertainty, and is computed by taking one combination of w and x at a time, determining the uncertainty of the y distribution, and then obtaining a weighted average for all of the w,x cells.

By transposing terms in Equation 1, we can define the multiple contingent uncertainty as

$$U(y\!:\!w,x) = U(y) - U_{wx}(y) \qquad [2]$$

The right hand side of this equation can then be rewritten to show the complete partitioning of the predictable uncertainty as

$$U(y\!:\!w,x) = U(y\!:\!w) + U(y\!:\!x) + U(y\!:\!\overline{wx}) \qquad [2a]$$

---

[2] For many of the equations presented here we shall not present calculational formulas or complete proofs, since these are available in at least a related form in other articles (especially **2**, **3**, **4**, **5**).

The three terms on the right of Equation 2a represent, in order, the uncertainty in y which is predictable from w, that which is predictable from x, and that which is preditable from unique combinations of w and x. The first two terms are simple contingent uncertainties, and are normally computed from collapsed two-variable matrices. The last term is an interaction uncertainty, and is computed either as a residual or as the difference between a simple contingent uncertainty and the same contingent uncertainty computed with the third variable held constant, rather than with the matrix collapsed over that variable. (See **3**, **4**.)

The partitioned uncertainties on the right of Equation 2a are completely analogous to partitioned variances in analysis of variance, and would in that case be called the main effect due to w, the main effect due to x, and the interaction between w and x.

Symmetric analysis. This approach to uncertainty analysis has the advantage of being completely analogous to the more familiar analysis of variance. It is possible, however, to approach uncertainty analysis from another point of view which allows us to deal with the total matrix of data as a unit. We can call this type a symmetric uncertainty analysis to distinguish it from the unidirectional analysis in which the uncertainty of just one variable is partitioned.

Since uncertainty analysis is nonmetric, we can look at the three-dimensional matrix shown in Fig. 1 as a single frequency distribution in which we have categories and proportions of total cases falling in each category. Thus we can compute the uncertainty of the total matrix of data,

$$U(w,x,y) = -\Sigma\{[p(i,j,k)] \times [\log_2 p(i,j,k)]\} \quad [3]$$

where $U(w,x,y)$ is the total uncertainty in the three-dimensional matrix, and $p(i,j,k)$ is as defined above. Now this total uncertainty can be broken down in many different ways, some of them being

$$\begin{aligned} U(w,x,y) & \\ &= U(y) + U_y(x) + U_{xy}(w) \quad [4a] \\ &= U(w) + U_w(x) + U_{wx}(y) \quad [4b] \\ &= U(w,x) + U_{wx}(y) \quad [4c] \\ &= U(w,y) + U_{wy}(x) \quad [4d] \\ &= U(x,y) + U_{xy}(w) \quad [4e] \end{aligned}$$

For these equations, the only terms needing additional explanation are $U_y(x)$ or equivalent forms, and $U(w,x)$ or equivalent forms. Any term of the form $U_y(x)$ is a conditional uncertainty, and is computed by taking one value at a time of the subscripted variable, determining the uncertainty of the variable in parentheses, then obtaining a weighted average over the subscripted variable. For such a calculation, the matrix is collapsed over the third variable. A term of the form $U(w,x)$ is a total uncertainty for a two-dimensional matrix, i.e., one collapsed over the third variable. It is computed from the $p(i,j)$, in this case, or the appropriate terms from any other combination of two variables.

$U(w,x,y)$ is a measure of the total uncertainty actually obtained in the matrix. We can also determine the maximum total uncertainty which could have been obtained from the matrix by determining what proportion of cases would fall in each cell if there were no contingencies between any of the variables. This ideal proportion, $P(i,j,k)$, is

$$P(i,j,k) = p(i)p(j)p(k) \quad [5]$$

in much the same way as is done for chi square. From these ideal proportions, representing those which would be obtained in a matrix with all variables completely independent, we can compute the maximum uncertainty in the matrix, $U_{\max}(w,x,y)$, as

$$U_{\max}(w,x,y) = -\Sigma\{[p(i)p(j)p(k)] \times [\log_2 p(i)p(j)p(k)]\} \quad [6]$$

This maximum uncertainty can itself be partitioned into components, and in this case the partitioning results in a simple set of terms.

$$U_{\max}(w,x,y) = U(w) + U(x) + U(y) \quad [7]$$

In other words, the maximum uncertainty which can be obtained in a matrix which is restricted only by the marginal totals is the sum of the uncertainties for the three variables taken one at a time.

Now if we have the maximum uncertainty which can be obtained in a matrix, and the uncertainty which actually is obtained, the difference between them must give us a measure of the total contingency or interrelatedness in the three-dimensional matrix. We can define a measure of this total contingency as

$$U(w:x:y) = U_{\max}(w,x,y) - U(w,x,y) \quad [8]$$

The term on the left is the total contingent uncertainty in the matrix, and the colon notation has been used as for other contingent uncertainty terms. It should be noted, however, that there is no implication of direction in the notation; the three symbols for the three variables can be written in any order.

The total contingent uncertainty can be partitioned also, and the use of Equation 7 with Equations 4c, 4d, or 4e leads to three different forms of the partitioning.

$$U(w:x:y)$$
$$= U(w:x) + U(y:w,x) \quad [9a]$$
$$= U(w:y) + U(x:w,y) \quad [9b]$$
$$= U(x:y) + U(w:x,y) \quad [9c]$$

All terms in these equations have been defined above, and need no further clarification here. These equations show that the total contingent uncertainty can be partitioned into a multiple contingent uncertainty in which one variable is predicted from the other two, plus a simple contingent uncertainty between the two variables used as predictors, and it makes no difference which variable is treated as a criterion and which as predictors for this statement to be true.

As in Equation 2a, each of the three multiple contingent uncertainties can itself be partitioned into components, so that the total contingent uncertainty can be partitioned more completely.

$$U(w:x:y) = U(w:x) + U(w:y) + U(x:y) + U(y:\overline{wx}) \quad [10a]$$
$$= U(w:x) + U(w:y) + U(x:y) + U(x:\overline{wy}) \quad [10b]$$
$$= U(w:x) + U(w:y) + U(x:y) + U(w:\overline{xy}) \quad [10c]$$

Thus the total contingent uncertainty consists of the three simple contingent uncertainties involving the three pairs of predictors, plus the interaction term. Although the three interaction terms have been written differently for each of the above equations, it is clear from these equations that all three interaction uncertainties are equal. In other words, there exists in the matrix all of the two-variable contingencies, plus a single interaction term which truly represents a three-dimensional pattern of inconsistencies. For this reason, it would be better to write the interaction uncertainty as

$U(\overline{wxy})$ to make clear the uniqueness of that term for the given matrix.

We can transpose the terms in Equation 9 to see the relation of each of the three possible multiple contingent uncertainties to the total contingent uncertainty, and we obtain

$$U(y\!:\!w,x) = U(w\!:\!x\!:\!y) - U(w\!:\!x) \quad [11a]$$

$$U(x\!:\!w,y) = U(w\!:\!x\!:\!y) - U(w\!:\!y) \quad [11b]$$

$$U(w\!:\!x,y) = U(w\!:\!x\!:\!y) - U(x\!:\!y) \quad [11c]$$

With the equations written in this form, some interesting relations between variables become clear. Each of these equations shows the multiple contingent uncertainty for predicting one of the three variables from the other two. The form of the equations is such that the amount of prediction available for each variable is equal to a constant less a variable term, which is in each case the contingent uncertainty between the two predictor variables. In other words, the amount of predictable uncertainty available for any variable in the matrix is the total contingent uncertainty less the contingent uncertainty between the two variables from which the predicting is being done. Therefore, the amount of prediction available in any direction in the matrix is inversely related to the contingency between the two predictors. For a given matrix the maximum possible predictability will be obtained when the predictor variables are orthogonal, i.e., have no contingent uncertainty between them. This seems intuitively sensible, and the uncertainty analysis presented here makes it clear that for any set of data this relation must hold.

In order to compute the exact amount of a multiple contingent uncertainty involving one criterion and two predictor variables, it is necessary to carry out calculations which involve the three-dimensional matrix. For many sets of data, such a calculation would be extremely laborious, if not prohibitive. Equation 11 shows that it is possible to determine how much better prediction is for one variable than for another without actually carrying out a complete three-dimensional calculation. The amount by which one of the three multiple contingent uncertainties is greater than another is exactly the same as the difference in the contingent uncertainties between the two sets of predictor variables, and these uncertainties can be calculated with the three-dimensional matrix collapsed to two dimensions. Actually, the interaction term is the one which requires a three-dimensional calculation to determine all the components of the multiple contingent uncertainty (see Equation 2a), and the essential fact which makes it possible to determine exact differences between multiple contingent uncertainties without the complete calculation is that the interaction uncertainty, as mentioned above, is the same for all three criterion variables. It is truly a term which applies to the complete matrix of data, and is not uniquely determined for each variable used as a criterion, as it would appear to be from a unidirectional partitioning approach to uncertainty analysis.

HIGHER-ORDER MATRICES

The entire discussion so far has been concerned with three-dimensional matrices. It is necessary to use at least three to show the nature of the principles, and it is easier to use just three to avoid undue complication. However, a brief discussion of at least the four-variable case will illustrate some additional factors. We noted above that it is necessary to have two-dimensional uncertainty calculations in order to determine the differences between the various three-dimensional, multiple contingent uncertainties. In other words, calculations with as many variables as there are in the matrix are necessary to obtain absolute numbers for the multiple contingent uncertainties, but calculations with one less variable suffice to determine the differences between the multiple terms. A similar situation exists for the general case, as is shown by the following equations, in which v is the fourth variable:

$$U(y{:}v,w,x) = U(v{:}w{:}x{:}y) - U(v{:}w{:}x) \qquad [12a]$$

$$U(x{:}v,w,y) = U(v{:}w{:}x{:}y) - U(v{:}w{:}y) \qquad [12b]$$

$$U(w{:}v,x,y) = U(v{:}w{:}x{:}y) - U(v{:}x{:}y) \qquad [12c]$$

$$U(v{:}w,x,y) = U(v{:}w{:}x{:}y) - U(w{:}x{:}y) \qquad [12d]$$

Each of the four multiple contingent uncertainties is the total contingent uncertainty in the four-variable matrix, $U(v{:}w{:}x{:}y)$, less the total contingent uncertainty between the three variables used as predictors, and exact differences between the multiples can be obtained with three-variable calculations. Similar conclusions concerning the variable whose prediction will be maximum can be made, since any variable which can be predicted from orthogonal predictor variables will have all of the contingent uncertainty in the matrix available for prediction of that variable.

Many situations exist, of course, where the data matrix has four or more variables, but where it is impractical or impossible to compute interaction terms involving even three variables. We noted above that when only three variables are involved, exact differences between the multiples can be determined because there is only one interaction term for the entire matrix. When four or more variables are involved, however, the interaction terms involved in each multiple contingent uncertainty are not identical. For example, with four variables, each multiple contingency requires four interaction terms, three involving three terms, and one involving all four terms. For any pair of multiple contingent uncertainties, three of these four interaction terms are identical, but the fourth will normally be different. As an example, the complete set of terms involved in two of the four multiple contingent uncertainties are

$$U(y{:}v,w,x) = U(y{:}v) + U(y{:}w) + U(y{:}x) + U(\overline{vwy}) + U(\overline{vxy}) + U(\overline{wxy}) + U(\overline{vwxy}) \qquad [13a]$$

$$U(x{:}v,w,y) = U(x{:}v) + U(x{:}w) + U(x{:}y) + U(\overline{vwx}) + U(\overline{vxy}) + U(\overline{wxy}) + U(\overline{vwxy}) \qquad [13b]$$

where $U(\overline{vwxy})$ is the four-variable interaction. Notice that only the terms $U(\overline{vwx})$ and $U(\overline{vwy})$ are different interactions for these two multiples, and that the other three interactions in each case are identical.

This fact means that estimated differences between the various multiple contingent uncertainties can be obtained with little error simply by adding up the two-variable contingent uncertainties involved in each case. The total difference in the interaction uncertainties must be small, since only one of four terms is different. All four interaction terms, of course, are needed to determine the absolute magnitude of the multiple contingent uncertainty, and, therefore, more error in estimating the absolute magnitude is likely in that case. When even more variables are included, the approximation by adding up the two-variable contingent uncertainties is more risky, since many more interaction terms are involved. However, it should be remembered (**3**, **4**, **5**) that interaction terms can take negative as well as postive values, so that it is often still reasonable to assume that the sum of all the interaction terms is negligible, even though any one of them may not be.

PSYCHOLOGICAL APPLICATIONS

The mathematical developments briefly described in the preceding sections have many interesting applications and implications in psychological research. In some cases, their primary value is to clarify the meaning of computations which are commonly carried out with the uncertainty measure. In other cases, these equations lead to some quite specific predictions about behavior which is related to nonrandom sequential events. A few of these implications will be discussed as illustrative cases.

Multivariate Information Transmission

One common application of information theory in psychology has been the use of information-transmission measures to determine the "channel capacity" of the human observer in transmitting, verbally or otherwise, information about a series of events or stimuli. Following the paradigm described by Garner and Hake (**2**), a typical experiment requires subjects to make an absolute response to a series of stimuli varying along some dimension, e.g., brightness or loudness, and the experimental result is expressed as an amount of information transmitted between stimuli and responses. If the information transmitted is less than the information contained in the stimuli and in the responses, it is assumed to represent a channel capacity for the observer. In the more general terminology of the present paper, information transmission is a two-variable contingent uncertainty.

Although many experiments have been done using this approach, for illustrative purposes we will refer to one by Garner (**1**). In this experiment, observers were asked to make an identifying response on a twenty-point scale to twenty different intensities of a tone. From the two-dimensional matrix involving just stimuli and responses, the contingent uncertainty between stimuli and responses is computed, and this figure can be interpreted as information transmitted. For this two-variable case, it makes no difference whether we state that information has been transmitted from stimuli to reponses or vice versa, or whether we talk about the amount of uncertainty in stimuli which can be predicted from responses, or vice versa. The measure of information transmission is completely bilateral, and can be given interpretation in either direction.

Additional computations were made in this experiment, however, to allow multiple predictions. For example, the multiple contingent uncertainty for predicting the response from knowledge of the observer and of the stimulus was calculated, and this multiple contingent uncertainty was, naturally, larger than the simple contingent uncertainty. It seems quite reasonable that if more variables are isolated for prediction purposes, more prediction should be available.

However, Equation **11** shows that, once more than one variable is used as a predictor, the bilateral interpretation of the contingent uncertainty is no longer possible. Let us look at this over-all situation as a three-variable matrix involving stimuli (S), responses (R), and observers (O). Now the multiple contingent uncertainty for predicting R from S and O is the total contingent uncertainty in the matrix minus the contingent uncertainty between S and O, the two predictor variables. In symbolic form,

$$U(R:S,O) = U(R:S:O) - U(S:O) \quad [14a]$$

In this experiment, however, as in most such experiments, all observers received the same simuli. Therefore, there is orthogonality between S and O, and the contingent uncertainty between them will be zero. Thus the contingent uncertainty available for predicting R is the same as the total contingent uncertainty in the three-dimensional matrix, and is at a maximum for this particular matrix.

Suppose now that we decide to predict stimuli from responses and observers—the type of prediction which is more realistic, since in the ordinary situation the response is transmitted, and from it and knowledge of the observer we make inferences or predictions about the stimulus. In this case, in symbolic form,

$$U(S:R,O) = U(R:S:O) - U(R:O) \quad [14b]$$

Again the multiple contingent uncertainty is the total contingent uncertainty in the matrix minus the contingent uncertainty between the two predictors, R and O. But in this case there is no reason why R and O should be orthogonal, and, in fact, if they are, there is little point to adding O as a predictor variable. Since $U(R:O)$ will be real and positive, prediction of S from R and O must be less than prediction of R from S and O, as in the former case.

These relations do not indicate that no improvement in prediction of S can be obtained by making use of knowledge about O, but only that prediction of S cannot be as great as prediction of R. If we consider only the prediction of S, then prediction from both R and from O can be broken down as follows:

$$U(S:R,O) = U(S:R) + U(S:O) + U(\overline{SRO}) \quad [15]$$

The third term on the right is the interaction. This equation shows the problem clearly, since no prediction of S can be obtained from just O as long as S and O are orthogonal. Additional prediction may, however, be obtained from the interaction term, which is to say that only inconsistencies in relations involving observers are useful in predicting S. For predicting R, however, both inconsistent and systematic differences by O in use of the R continuum provide additional prediction.

A simple example will illustrate these relations. Suppose that two observers rate two stimuli with perfect consistency, but one observer uses the numbers "1" and "2", while the other uses the numbers "3" and "4", for the first and second stimuli, respectively. Now if we know the response we can perfectly predict the stimulus, and it is no help to us to know which observer uses which responses. If we use the stimulus to predict the response, however, we must first know which observer is involved to know which response is used. On the other hand, suppose that both observers use the numbers "1" and "2", but in reverse relation to the two stimuli. Knowing the stimulus alone is now of no value in predicting the reponse, or vice versa.

However, if knowledge of the observer is added, then perfect prediction of the response from the stimulus or of the stimulus from the response is possible, due to the fact that the addition of observers as a predictor variable adds an interaction term in which all of the available prediction lies.

As a concrete example, these two cases are extreme and serve only to illustrate the principles. One fact remains, however, which is that, in such a case, prediction of stimuli can never be better than prediction of responses, and will normally be somewhat poorer.

This type of problem can be carried even further. In the same experiment, Garner actually determined the multiple contingent uncertainty when three predictor variables were used, the fourth variable being the stimulus which preceded the one being judged (P). Furthermore, in the experimental design, all pairs of stimuli were made to occur equally often, so that P and S were orthogonally related. Now the two types of prediction are

$$U(R{:}S,O,P) = U(R{:}S{:}O{:}P) - U(S{:}O{:}P) \quad [16a]$$

and

$$U(S{:}R,O,P) = U(R{:}S{:}O{:}P) - U(R{:}O{:}P) \quad [16b]$$

following the form of Equation 12. In this case, when R is predicted from S, O, and P, the multiple contingent uncertainty is the total contingent uncertainty in the four-variable matrix minus the total contingent uncertainty in the three-variable matrix involving the three predictor variables. Since these three predictor variables were all orthogonal in the experimental design, this term is zero, and this multiple contingent uncertainty contains all of the contingent uncertainty available in the matrix.

When, however, S is predicted from the other three variables, the orthogonal condition does not hold, since of the three variables only O and P are orthogonal. This term will thus be positive, and will be fairly substantial, since it includes contingent uncertainties between R and O, and between R and P, as well as the three-variable interaction which must (in this case) also be positive. Thus, there would be substantial error in assuming that these two multiple contingent uncertainties are the same.

Redundancy of Printed English

One common application of information theory and uncertainty analysis is to situations where events occur in a sequence which is not completely random, i.e., in which there are sequential dependencies. Much of human behavior is concerned with making predictions about events which are unknown, or only partially known, on the basis of sequentially provided information. Language is a very good example of such an event series in which there exist sequential dependencies, and several studies have been done on printed English to determine how much predictability is available from sequences of letters. The term "redundancy" has been used to describe the amount of uncertainty in a given letter which is predictable on the basis of other known letters. The equations described above point out some interesting aspects of redundancy of printed English.

When we do an analysis of sequences of letters, we form a data matrix in which the series of letters itself is one variable, and the series displaced by one or more units constitute the other variables. For present purposes we shall refer to the first order series as variable w, the series displaced one step as variable x, and the series displaced two steps as variable y, and we will be dealing with a three-dimensional matrix. In this special three-variable case there are some restrictions on the contingent uncertainties which exist. For example, we know that the term $U(w:x)$ is identical to the term $U(x:y)$, since in both cases we are describing the contingent uncertainty for a one-step displacement of the series. In addition, the contingent uncertainty $U(w:y)$ is less than either of the others, since that is the contingency for a two-step displacement, and a letter cannot be predicted as well from one two steps behind as from the immediately preceding one.

Now we know from Equation 11 that the multiple contingent uncertainty for predicting x, the middle letter of a series of three, must be greater than that for predicting either w or y from the other two, since when x is being predicted, the contingent uncertainty between the two predictor variables, w and y, is less than the contingent uncertainties between the other pairs of predictor variables. In other words, prediction of the middle letter of a sequence of three must be better than prediction of either end of the sequence on the basis of the other two. Similar relations hold for longer sequences of letters, and we can make the general statement that if any fixed number of letters is available for prediction, the best prediction will occur if the letter being predicted is in the middle of the total sequence rather than at either end. Miller and Friedman (6) have experimentally shown this to be true for sequences of 5, 7, and 11 letters.

Newman and Gerstman (7) have provided data which allow us to see just how much better prediction of middle letters should be, compared to prediction of end letters. Their data show, for the three-letter case, that the contingent uncertainty for a one-step displacement is 0.91 bit, and for a two-step displacement it is 0.42 bit. This difference is 0.49 bit, and reference to Equation 11 shows that this is the exact amount by which prediction of the middle of three letters will be better than prediction of either end. This is a substantial difference.

As pointed out above, the situation becomes more complicated when we deal with longer letter sequences because of the greater number of interaction terms. However, we can assume that these terms are negligible, and determine the multiple contingent uncertainty for predicting omitted letters from other length sequences simply by adding up the various two-term contingent uncertainties involved in the multiple contingent uncertainty. Using again the Newman and Gerstman data, for an eleven-letter sequence, the contingent uncertainty for predicting either end letter on the basis of the other ten is 2.17 bits, while that for predicting the middle letter, on the basis of five letters on each side, is 3.72 bits. For seven-letter sequences, prediction of the end letters gives a multiple contingent uncertainty of 1.95 bits, while prediction of the middle letter gives 3.18 bits. These figures compare quite well with those found by Miller and Friedman, although the agreement is better for middle letters than for ends. It seems, from their data, that humans make more effective use of information presented symmetrically than they do of unilateral information, since their data show higher residual uncertainties for the ends than these calculations predict.

These relations point out the need

for a more accurate interpretation of the meaning of the term redundancy. This term has usually been defined for prediction of letters on the basis of preceding letters only, and the figure of 50 per cent redundancy is accepted as a reasonable approximation. Such a figure, however, can be interpreted quantitatively and accurately only when it is applied to this particular prediction situation. We cannot infer from it how many letters in printed English can be omitted and still allow perfect reconstruction. Clearly, with a figure of 50 per cent redundancy for one direction only, we cannot expect perfect reproducibility unless we have this amount available on both sides of the item being predicted. In other words, we need 100 per cent redundancy for perfect prediction, and this can be obtained only when a single letter is omitted from a very long sequence, so that maximum redundancy is available on both sides of the letter being predicted. Perhaps our thinking on this problem has been limited by our natural tendency to deal with antecedent and consequent events; but we can think of both antecedent and subsequent events as determining the consequent events, and combinations of these certainly are more effective than either alone.

Response to a Continuing Series

A type of behavior for which the relations shown here have important implications is that in which a continuing series of events occurs and a human must make a differential response to each event as it occurs. Such tasks as copying telegraphic code, taking dictation, reporting target positions on a radar, and target tracking are all of this type.

To talk about a simplified task of this sort, let us assume that an operator is read a series of numbers and he is required to push a different button for each different number he hears, much as he would if he were adding up a column of figures on a machine as they were being read to him. The operator, of course, is trying to carry out his task with minimum error, so that he wants to get all the information possible about what number was read to him. Initially, let us suppose also that the numbers do not occur in a completely random order, but that each number is to some extent predictable from the preceding ones in a fairly simple way. Such a series, of course, will show sequential dependencies, and, with rare exceptions, the contingent uncertainties between the series and its displacements will be largest for adjacent numbers and will decrease as the number series is displaced from itself more and more. Now we know from the relations shown in this paper that the maximum predictability for any single number will be obtained if the operator is able to use symmetrical prediction of it, i.e., to use both forward and backward prediction. In order to use both forward and backward prediction, however, he must delay his response to a particular number until he has heard some subsequent numbers. He must, in other words, lag behind the presented series. This is, of course, a phenomenon commonly observed in everyday life.

Suppose we want to determine the optimum lag which such an operator should use. We can make some reasonable predictions, but these predictions will depend on a number of factors.

1. *The memory function.* It is clear that the operator cannot delay his response so long that he cannot remember all of the items which are potentially useful for prediction of any one

TABLE 1

ILLUSTRATION OF EFFECT OF MEMORY AND CONTINGENT UNCERTAINTIES ON OPTIMUM LAG

Unit Step	Contingent Uncertainty	Lag of 2		Lag of 3		Lag of 1	
		Weight	Weighted Value	Weight	Weighted Value	Weight	Weighted Value
−3	.30					.2	.06
−2	.50	.2	.10			.4	.20
−1	1.00	.4	.40	.2	.20	.6	.60
0		.6		.4		.8	
+1	1.00	.8	.80	.6	.60	1.0	1.00
+2	.50	1.0	.50	.8	.40		
+3	.30			1.0	.30		
Σ weighted values			1.80		1.50		1.86

item. Let us assume that a reasonable memory span for such a task is five units. Now if all five units can be remembered equally well, the optimum lag for the subject would be two units behind the one to which the response is being made, since this lag will give two units on either side of the one in question, with the optimum of completely symmetrical prediction.

Actually, it is unrealistic to assume that all of the items can be remembered equally well. It seems more reasonable to assume that the items last heard will be more clearly remembered, and thus that the contingent uncertainty available from them will be more useful. We can illustrate the effects of such a memory function by using a weighting function which gives greater value to recent items than to remote items. Table 1 contains a set of purely hypothetical numbers to illustrate how such a memory factor would affect the optimum lag. The first column shows the unit step for each item, with "O" being the item to which the response is to be made, negative numbers referring to items which occurred before that one, and positive numbers being the later, more recent items. Column 2 shows the contingent uncertainty for each displacement with respect to the "O" item. Column 3 shows the weights which have been assigned to each item being used for predictive purposes, with the largest weight for the most recent item. Column 4 shows the resultant weighted value of the contingent uncertainty for each step, with the total at the bottom. For these two columns, we have assumed a lag of two steps. The next two columns show the same calculations for a lag of three steps, and the last two columns for a lag of one step.

With an equiweighted function, the maximum multiple contingent uncertainty occurs with a lag of two, and has a value of 3.00 bits (i.e., all weights are 1.0). When the differential weights are used, however, the maximum-weighted contingent uncertainty occurs with a lag of one, and has a value of 1.86 bits. Thus, if we assume a memory differential for the various steps, two things happen: the total contingent uncertainty decreases, and the optimum lag also decreases. In general, where responses are made to a continued time series, we would expect the optimum lag to be somewhat less than that which gives complete symmetry to the predictors.

2. *The amount of redundancy.* A second factor which affects the optimum lag is the total amount of redundancy in the series, as well as its distribution over the various steps. This point need not be labored. A lag which gains no predictability can be of no value. Clearly if the series is completely random (no redundancy), there is no advantage to the operator in lagging because he gains no redundancy or information when he does. Similarly, if only one item on either side gives any predictability, then no lag beyond one step is ever of any value.

3. *The amount of noise.* One last factor which will affect the optimum lag for the operator is the amount of noise whose effects have to be offset. In our illustration, the term noise can refer to acoustical noise, or in the more general case, it can refer to any factor which decreases the probability of a correct perception of the item. The importance of noise is that it determines the extent to which the operator needs the predictability which he can obtain from the series redundancy. Suppose, for example, that the task is being carried out under ideal circumstances, where each number can be heard quite clearly. Under these circumstances, the operator can respond immediately with essentially perfect accuracy, and it would make little difference to him whether the series is partially redundant or not. In either case he would not lag, since there is nothing to be gained by it. If, on the other hand, there is sufficient noise to make the perception imperfect, then the operator will frequently find himself unsure of the proper response, and will need some additional redundancy to improve the probability of his making a correct response.

Thus, in summary, we can assume that with this type of task operators will lag if there is some noise in the situation and if the series has sequential dependencies in it. If these conditions are satisfied, then the exact amount of the lag will depend on his memory for the items presented and on the distribution of the redundancy in the series of events.

SUMMARY

In this paper we have presented some equations originating in information theory which show some valuable properties of uncertainty analysis. These equations are presented in a form which emphasizes the fact that an uncertainty analysis can be carried out so that all variables involved can be dealt with in a completely symmetric manner. It is possible, for example, to talk about a single number which represents the total contingent uncertainty in a multidimensional matrix, and then to demonstrate how parts or all of this contingent uncertainty can be made available for prediction of just one of the variables. From such equations it is possible to demonstrate certain relations which must hold between prediction of different variables in a single matrix.

These relations have important implications for several areas of psychology, either in clarifying the meaning of terms used in statistical analysis or in showing how behavior ought to occur if it is to maximize the use of information. As one illustration, it is shown that information transmission in the multidimensional case cannot be as good when stimuli are predicted as when responses are predicted. Another illustration shows that the usual measure of redundancy of printed English can be interpreted accurately only when prediction is

made unilaterally, but that maximum predictability of single letters occurs when prediction (on the basis of redundancy) is allowed to operate in both the forward and backward directions. A third illustration shows that when human operators have to respond to a continuing series which is nonrandom, optimum performance (with an error criterion) will occur only when the operator lags behind the series to which he is responding; furthermore, the optimum lag will be a function of the memory of the operator, the amount of noise in the situation, and the distribution of redundancy throughout the series.

These examples are in no sense inclusive, and the implications of equations presented here range into many areas of psychology. These illustrations are sufficient, however, to point out several possible directions for research, and to clarify some concepts now in use in psychology.

REFERENCES

1. GARNER, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46**, 373–380.
2. GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
3. GARNER, W. R., & MCGILL, W. J. The relation between information and variance analyses. *Psychometrika*, 1956, **21**, 219–228.
4. MCGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.
5. MCGILL, W. J., & QUASTLER, H. Standardized nomenclature. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955. Pp. 83–92.
6. MILLER, G. A., & FRIEDMAN, E. A. The reconstruction of mutilated English texts. *Information and Control*, 1957, **1**, 38–55.
7. NEWMAN, E. B., & GERSTMAN, L. J. A new method for analyzing printed English. *J. exp. Psychol.*, 1952, **44**, 114–125.
8. SHANNON, C. E. A mathematical theory of communication. *Bell Syst. tech. J.*, 1948, **27**, 379–423, 623–656.

(Received February 27, 1958)

Psychological Review
Vol. 65, No. 4, 1958

HEREDITY, ENVIRONMENT, AND THE QUESTION "HOW?"[1]

ANNE ANASTASI

Fordham University

Two or three decades ago, the so-called heredity-environment question was the center of lively controversy. Today, on the other hand, many psychologists look upon it as a dead issue. It is now generally conceded that both hereditary and environmental factors enter into all behavior. The reacting organism is a product of its genes and its past environment, while present environment provides the immediate stimulus for current behavior. To be sure, it can be argued that, although a given trait may result from the combined influence of hereditary and environmental factors, a specific difference in this trait between individuals or between groups may be traceable to either hereditary or environmental factors alone. The design of most traditional investigations undertaken to identify such factors, however, has been such as to yield inconclusive answers. The same set of data has frequently led to opposite conclusions in the hands of psychologists with different orientations.

Nor have efforts to determine the proportional contribution of hereditary and environmental factors to observed individual differences in given traits met with any greater success. Apart from difficulties in controlling conditions, such investigations have usually been based upon the implicit assumption that hereditary and environmental factors combine in an additive fashion. Both geneticists and psychologists have repeatedly demonstrated, however, that a more tenable hypothesis is that of interaction (15, 22, 28, 40). In other words, the nature and extent of the influence of each type of factor depend upon the contribution of the other. Thus the proportional contribution of heredity to the variance of a given trait, rather than being a constant, will vary under different environmental conditions. Similarly, under different hereditary conditions, the relative contribution of environment will differ. Studies designed to estimate the proportional contribution of heredity and environment, however, have rarely included measures of such interaction. The only possible conclusion from such research would thus seem to be that both heredity and environment contribute to all behavior traits and that the extent of their respective contributions cannot be specified for any trait. Small wonder that some psychologists regard the heredity-environment question as unworthy of further consideration!

But is this really all we can find out about the operation of heredity and environment in the etiology of behavior? Perhaps we have simply been asking the wrong questions. The traditional questions about heredity and environment may be intrinsically unanswerable. Psychologists began by asking *which* type of factor, hereditary or environmental, is responsible for individual differences in a given trait. Later, they tried to discover *how much* of the variance was attributable to heredity and how much to environment. It is the primary contention of this paper that a more fruitful approach is to be found in the question *"How?"* There is still much to be learned about the specific *modus operandi* of hereditary and environmental factors in the development of behavioral differences. And there are several current lines of research which offer promising techniques for answering the question "How?"

---

[1] Address of the President, Division of General Psychology, American Psychological Association, September 4, 1957.

VARIETY OF INTERACTION MECHANISMS

Hereditary factors. If we examine some of the specific ways in which hereditary factors may influence behavior, we cannot fail but be impressed by their wide diversity. At one extreme, we find such conditions as phenylpyruvic amentia and amaurotic idiocy. In these cases, certain essential physical prerequisites for normal intellectual development are lacking as a result of hereditary metabolic disorders. In our present state of knowledge, there is no environmental factor which can completely counteract this hereditary deficit. The individual will be mentally defective, regardless of the type of environmental conditions under which he is reared.

A somewhat different situation is illustrated by hereditary deafness, which may lead to intellectual retardation through interference with normal social interaction, language development, and schooling. In such a case, however, the hereditary handicap can be offset by appropriate adaptations of training procedures. It has been said, in fact, that the degree of intellectual backwardness of the deaf is an index of the state of development of special instructional facilities. As the latter improve, the intellectual retardation associated with deafness is correspondingly reduced.

A third example is provided by inherited susceptibility to certain physical diseases, with consequent protracted ill health. If environmental conditions are such that illness does in fact develop, a number of different behavioral effects may follow. Intellectually, the individual may be handicapped by his inability to attend school regularly. On the other hand, depending upon age of onset, home conditions, parental status, and similar factors, poor health may have the effect of concentrating the individual's energies upon intellectual pursuits. The curtailment of participation in athletics and social functions may serve to strengthen interest in reading and other sedentary activities. Concomitant circumstances would also determine the influence of such illness upon personality development. And it is well known that the latter effects could run the gamut from a deepening of human sympathy to psychiatric breakdown.

Finally, heredity may influence behavior through the mechanism of social stereotypes. A wide variety of inherited physical characteristics have served as the visible cues for identifying such stereotypes. These cues thus lead to behavioral restrictions or opportunities and—at a more subtle level—to social attitudes and expectancies. The individual's own self concept tends gradually to reflect such expectancies. All of these influences eventually leave their mark upon his abilities and inabilities, his emotional reactions, goals, ambitions, and outlook on life.

The geneticist Dobzhansky illustrates this type of mechanism by means of a dramatic hypothetical situation. He points out that, if there were a culture in which the carriers of blood group AB were considered aristocrats and those of blood group O laborers, then the blood-group genes would become important hereditary determiners of behavior (12, p. 147). Obviously the association between blood group and behavior would be specific to that culture. But such specificity is an essential property of the causal mechanism under consideration.

More realistic examples are not hard to find. The most familiar instances occur in connection with constitutional types, sex, and race. Sex and skin pigmentation obviously depend upon heredity. General body build is strongly influenced by hereditary components, although also susceptible to environmental modification. That all these physical characteristics may exert a pronounced effect upon behavior within a given culture is well known. It is equally apparent, of course, that in different cultures the behavioral correlates of such hereditary physical traits may be quite unlike. A specific physical cue may be completely unrelated to individual differences in psychological traits in one culture, while closely correlated with them in another. Or it may be associated with totally dissimilar behavior characteristics in two different cultures.

It might be objected that some of the illustrations which have been cited do not properly exemplify the operation of hereditary mechanisms in behavior development, since hereditary factors enter only indirectly into the behavior in question. Closer examination, however, shows this distinction to be untenable. First it may be noted that the influence of heredity upon behavior is always indirect. No psychological trait is ever inherited as such. All we can ever say directly from behavioral observations is that a given trait shows evidence of being influenced by certain "inheritable unknowns." This merely defines a problem for genetic research; it does not provide a causal explanation. Unlike the blood groups, which are close to the level of primary gene products, psychological traits are related to genes by highly indirect and devious routes. Even the mental deficiency associated with phenylketonuria is several steps removed from the chemically defective genes that represent its hereditary basis. Moreover, hereditary influences cannot be dichotomized into the more direct and the less direct. Rather do they represent a whole "continuum of indirectness," along which are found all degrees of remoteness of causal links. The examples already cited illustrate a few of the points on this continuum.

It should be noted that as we proceed along the continuum of indirectness, the range of variation of possible outcomes of hereditary factors expands rapidly. At each step in the causal chain, there is fresh opportunity for interaction with other hereditary factors as well as with environmental factors. And since each interaction in turn determines the direction of subsequent interactions, there is an ever-widening network of possible outcomes. If we visualize a simple sequential grid with only two alternatives at each point, it is obvious that there are two possible outcomes in the one-stage situation, four outcomes at the second stage, eight at the third, and so on in geometric progression. The actual situation is undoubtedly much more complex, since there will usually be more than two alternatives at any one point.

In the case of the blood groups, the relation to specific genes is so close that no other concomitant hereditary or environmental conditions can alter the outcome. If the organism survives at all, it will have the blood group determined by its genes. Among psychological traits, on the other hand, some variation in outcome is always possible as a result of concurrent circumstances. Even in cases of phenylketonuria, intellectual development will exhibit some relationship with the type of care and training available to the individual. That behavioral outcomes show progressive diversification as we proceed along the continuum of indirectness is brought out by the other examples which were cited. Chronic illness *can* lead to scholarly renown or to intellectual immaturity; a mesomorphic physique *can* be a contributing factor in juvenile delinquency or in the attainment of a college presidency! Published data on Sheldon somatotypes provide some support for both of the latter outcomes.

Parenthetically, it may be noted that geneticists have sometimes used the term "norm of reaction" to designate the range of variation of possible outcomes of gene properties (cf. 13, p. 161). Thus heredity sets the "norm" or limits within which environmental differences determine the eventual outcome. In the case of some traits, such as blood groups or eye color, this norm is much narrower than in the case of other traits. Owing to the rather different psychological connotations of both the words "norm" and "reaction," however, it seems less confusing to speak of the "range of variation" in this context.

A large portion of the continuum of hereditary influences which we have described coincides with the domain of somatopsychological relations, as defined by Barker et al. (6). Under this heading, Barker includes "variations in physique that affect the psychological situation of a person by influencing the effectiveness of his body as a tool for actions or by serving as a stimulus to himself or others" (6, p. 1). Relatively direct neurological influences on behavior, which have been the traditional concern of physiological psychology, are excluded from this definition, Barker being primarily concerned with what he calls the "social psychology of physique." Of the examples cited in the present paper, deafness, severe illness, and the physical characteristics associated with social stereotypes would meet the specifications of somatopsychological factors.

The somatic factors to which Barker refers, however, are not limited to those of hereditary origin. Bodily conditions attributable to environmental causes operate in the same sorts of somatopsychological relations as those traceable to heredity. In fact, heredity-environment distinctions play a minor part in Barker's approach.

Environmental factors: organic. Turning now to an analysis of the role of environmental factors in behavior, we find the same etiological mechanisms which were observed in the case of hereditary factors. First, however, we must differentiate between two classes of environmental influences: (*a*) those producing organic effects which may in turn influence behavior and (*b*) those serving as direct stimuli for psychological reactions. The former may be illustrated by food intake or by exposure to bacterial infection; the latter, by tribal initiation ceremonies or by a course in algebra. There are no completely satisfactory names by which to designate these two classes of influences. In an earlier paper by Anastasi and Foley (4), the terms "structural" and "functional" were employed. However, "organic" and "behavioral" have the advantage of greater familiarity in this context and may be less open to misinterpretation. Accordingly, these terms will be used in the present paper.

Like hereditary factors, environmental influences of an organic nature can also be ordered along a continuum of indirectness with regard to their relation to behavior. This continuum closely parallels that of hereditary factors. One end is typified by such conditions as mental deficiency resulting from cerebral birth injury or from prenatal nutritional inadequacies. A more indirect etiological mechanism is illustrated by severe motor disorder—as in certain cases of cerebral palsy—*without* accompanying injury to higher neurological centers. In such instances, intellectual retardation may occur as an indirect result of the motor handicap, through the curtailment of educational and social activities. Obviously this causal mechanism corresponds closely to that of hereditary deafness cited earlier in the paper.

Finally, we may consider an environmental parallel to the previously discussed social stereotypes which were mediated by hereditary physical cues. Let us suppose that a young woman with mousy brown hair becomes transformed into a dazzling golden blonde through environmental techniques currently available in our culture. It is highly probable that this metamorphosis will alter, not only the reactions of her associates toward her, but also her own self concept and subsequent behavior. The effects could range all the way from a rise in social poise to a drop in clerical accuracy!

Among the examples of environmentally determined organic influences which have been described, all but the first two fit Barker's definition of somatopsychological factors. With the exception of birth injuries and nutritional deficiencies, all fall within the social psychology of physique. Nevertheless, the individual factors exhibit wide diversity in their specific *modus operandi*—a diversity which has important practical as well as theoretical implications.

Environmental factors: behavioral. The second major class of environmental factors—the behavioral as contrasted to the organic—are by definition direct influences. The immediate effect of such environmental factors is always a behavioral change. To be sure, some of the initial behavioral effects may themselves indirectly affect the individual's later behavior. But this relationship can perhaps be best conceptualized in terms of breadth and permanence of effects. Thus it could be said that we are now dealing, not with a continuum of indirectness, as in the case of hereditary and organic-environmental factors, but rather with a continuum of breadth.

Social class membership may serve as an illustration of a relatively broad, pervasive, and enduring environmental factor. Its influence upon behavior development may operate through many channels. Thus social level may determine the range and nature of intellectual stimulation provided by home and community through books, music, art, play activities, and the like. Even more far-reaching may be the effects upon interests and motivation, as illustrated by the desire to perform abstract intellectual tasks, to surpass others in competitive situations, to succeed in school, or to gain social approval. Emotional and social traits may likewise be influenced by the nature of interpersonal relations characterizing homes at different socioeconomic levels. Somewhat more restricted in scope than social class, although still exerting a relatively broad influence, is amount of formal schooling which the individual is able to obtain.

A factor which may be wide or narrow in its effects, depending upon concomitant circumstances, is language handicap. Thus the bilingualism of an adult who moves to a foreign country with inadequate mastery of the new language represents a relatively limited handicap which can be readily overcome in most cases. At most, the difficulty is one of communication. On the other hand, some kinds of bilingualism in childhood may exert a retarding influence upon intellectual development and may under certain conditions affect personality development adversely (2, 5, 10). A common pattern in the homes of immigrants is that the child speaks one language at home and another in school, so that his knowledge of each language is limited to certain types of situations. Inadequate facility with the language of the school interferes with the acquisition of basic concepts, intellectual skills, and information. The frustration engendered by scholastic difficulties may in turn lead to discouragement and general dislike of school. Such reactions can be found, for example, among a number of Puerto Rican children in New York City schools (3). In the case of certain groups, moreover, the child's foreign language background may be perceived by himself and his associates as a symbol of minority group status and may thereby augment any emotional maladjustment arising from such status (34).

A highly restricted environmental influence is to be found in the opportunity to acquire specific items of information occurring in a particular intelligence test. The fact that such opportunities may vary with culture, social class, or individual experiential background is at the basis of the test user's concern with the problem of coaching and with "culture-free" or "culture-fair" tests (cf. 1, 2). If the advantage or disadvantage which such experiential differences confer upon certain individuals is strictly confined to performance on the given test, it will obviously reduce the validity of the test and should be eliminated.

In this connection, however, it is essential to know the breadth of the environmental influence in question. A fallacy inherent in many attempts to develop culture-fair tests is that the breadth of cultural differentials is not taken into account. Failure to consider breadth of effect likewise characterizes certain discussions of coaching. If, in coaching a student for a college admission test, we can improve his knowledge of verbal concepts and his reading comprehension, he will be better equipped to succeed in college courses. His performance level will thus be raised, not only on the test, but also on the criterion which the test is intended to predict. To try to devise a test which is not susceptible to such coaching would merely reduce the effectiveness of the test. Similarly, efforts to rule out cultural differentials from test items so as to make them equally "fair" to subjects in different social classes or in different cultures may merely limit the usefulness of the test, since the same cultural differentials may operate within the broader area of behavior which the test is designed to sample.

METHODOLOGICAL APPROACHES

The examples considered so far should suffice to highlight the wide variety of ways in which hereditary and environmental factors may interact in the course of behavior development. There is clearly a need for identifying explicitly the etiological mechanism whereby any given hereditary or environmental condition ultimately leads to a behavioral characteristic—in other words, the "how" of heredity and environment. Accordingly, we may now take a quick look at some promising methodological approaches to the question "how."

Within the past decade, an increasing number of studies have been designed to trace the connection between specific factors in the hereditary backgrounds or in the reactional biographies of individuals and their observed behavioral characteristics. There has been a definite shift away from the predominantly descriptive and correlational approach of the earlier decades toward more deliberate attempts to verify explanatory hypotheses. Similarly, the cataloguing of group differences in psychological traits has been giving way gradually to research on *changes* in group characteristics following altered conditions.

Among recent methodological developments, we have chosen seven as being particularly relevant to the analysis of etiological mechanisms. The first represents an extension of selective breeding investigations to permit the identification of specific hereditary conditions underlying the observed behavioral differences. When early selective breeding investigations such as those of Tryon

(36) on rats indicated that "maze learning ability" was inherited, we were still a long way from knowing what was actually being transmitted by the genes. It was obviously not "maze learning ability" as such. Twenty—or even ten—years ago, some psychologists would have suggested that it was probably general intelligence. And a few might even have drawn a parallel with the inheritance of human intelligence.

But today investigators have been asking: Just what makes one group of rats learn mazes more quickly than the other? Is it differences in motivation, emotionality, speed of running, general activity level? If so, are these behavioral characteristics in turn dependent upon group differences in glandular development, body weight, brain size, biochemical factors, or some other organic conditions? A number of recent and ongoing investigations indicate that attempts are being made to trace, at least part of the way, the steps whereby certain chemical properties of the genes may ultimately lead to specified behavior characteristics.

An example of such a study is provided by Searle's (31) follow-up of Tryon's research. Working with the strains of maze-bright and maze-dull rats developed by Tryon, Searle demonstrated that the two strains differed in a number of emotional and motivational factors, rather than in ability. Thus the strain differences were traced one step further, although many links still remain to be found between maze learning and genes. A promising methodological development within the same general area is to be found in the recent research of Hirsch and Tryon (18). Utilizing a specially devised technique for measuring individual differences in behavior among lower organisms, these investigators launched a series of studies on selective breeding for behavioral characteristics in the fruit fly, *Drosophila.* Such research can capitalize on the mass of available genetic knowledge regarding the morphology of *Drosophila,* as well as on other advantages of using such an organism in genetic studies.

Further evidence of current interest in the specific hereditary factors which influence behavior is to be found in an extensive research program in progress at the Jackson Memorial Laboratory, under the direction of Scott and Fuller (30). In general, the project is concerned with the behavioral characteristics of various breeds and cross-breeds of dogs. Analyses of some of the data gathered to date again suggest that "differences in performance are produced by differences in emotional, motivational, and peripheral processes, and that genetically caused differences in central processes may be either slight or non-existent" (29, p. 225). In other parts of the same project, breed differences in physiological characteristics, which may in turn be related to behavioral differences, have been established.

A second line of attack is the exploration of possible relationships between behavioral characteristics and physiological variables which may in turn be traceable to hereditary factors. Research on EEG, autonomic balance, metabolic processes, and biochemical factors illustrates this approach. A lucid demonstration of the process of tracing a psychological condition to genetic factors is provided by the identification and subsequent investigation of phenylpyruvic amentia. In this case, the causal chain from defective gene, through metabolic disorder and consequent cerebral malfunctioning, to feeblemindedness and other overt symptoms can be described step by step (cf. 32; 33, pp. 389–391). Also relevant are the recent researches on neurological and biochemical correlates of schizophrenia (9). Owing to inadequate methodological controls, however, most of the findings of the latter studies must be regarded as tentative (19).

Prenatal environmental factors provide a third avenue of fruitful investigation. Especially noteworthy is the recent work of Pasamanick and his associates (27), which demonstrated a tie-up between socioeconomic level, complications of pregnancy and parturition, and psychological disorders of the offspring. In a series of studies on large samples of whites and Negroes in Baltimore, these investigators showed that various prenatal and paranatal disorders are significantly related to the occurrence of mental defect and psychiatric disorders in the child. An important source of such irregularities in the process of childbearing and birth is to be found in deficiencies of maternal diet and in other conditions associated with low socioeconomic status. An analysis of the data did in fact reveal a much higher frequency of all such medical complications in lower than in higher socioeconomic levels, and a higher frequency among Negroes than among whites.

Direct evidence of the influence of prenatal nutritional factors upon subsequent intellectual development is to be found in a recent, well controlled experiment by Harrell et al. (16). The subjects were pregnant women in low-income groups, whose normal diets were generally quite deficient. A dietary supplement was administered to some of these women during pregnancy and lactation, while an equated control group received placebos. When tested at the ages of three and four years, the offspring of the experimental group obtained a significantly higher mean IQ than did the offspring of the controls.

Mention should also be made of animal experiments on the effects of such factors as prenatal radiation and neonatal asphyxia upon cerebral anomalies as well as upon subsequent behavior development. These experimental studies merge imperceptibly into the fourth approach to be considered, namely, the investigation of the influence of early experience upon the eventual behavioral characteristics of animals. Research in this area has been accumulating at a rapid rate. In 1954, Beach and Jaynes (8) surveyed this literature for the *Psychological Bulletin,* listing over 130 references. Several new studies have appeared since that date (e.g., 14, 21, 24, 25, 35). The variety of factors covered ranges from the type and quantity of available food to the extent of contact with human culture. A large number of experiments have been concerned with various forms of sensory deprivation and with diminished opportunities for motor exercise. Effects have been observed in many kinds of animals and in almost all aspects of behavior, including perceptual responses, motor activity, learning, emotionality, and social reactions.

In their review, Beach and Jaynes pointed out that research in this area has been stimulated by at least four distinct theoretical interests. Some studies were motivated by the traditional concern with the relative contribution of maturation and learning to behavior development. Others were designed in an effort to test certain psychoanalytic theories regarding infantile experiences, as illustrated by studies which limited the feeding responses of young animals. A third relevant influence is to be found in the work of the European biologist Lorenz (23) on early social stimulation of birds, and in particular on the special type of learning for which the term "imprinting" has been coined. A relatively large number of recent studies have centered around Hebb's (17) theory regarding the importance of early perceptual experiences upon subsequent performance in learning situations. All this research represents a rapidly growing and promising attack on the *modus operandi* of specific environmental factors.

The human counterpart of these animal studies may be found in the comparative investigation of child-rearing practices in different cultures and subcultures. This represents the fifth approach in our list. An outstanding example of such a study is that by Whiting and Child (38), published in 1953. Utilizing data on 75 primitive societies from the Cross-Cultural Files of the Yale Institute of Human Relations, these investigators set out to test a number of hypotheses regarding the relationships between child-rearing practices and personality development. This analysis was followed up by field observations in five cultures, the results of which have not yet been reported (cf. 37).

Within our own culture, similar surveys have been concerned with the diverse psychological environments provided by different social classes (11). Of particular interest are the study by Williams and Scott (39) on the association between socioeconomic level, permissiveness, and motor development among Negro children, and the exploratory research by Milner (26) on the relationship between reading readiness in first-grade children and patterns of parent-child interaction. Milner found that upon school entrance the lower-class child seems to lack chiefly two advantages enjoyed by the middle-class child. The first is described as "a warm positive family atmosphere or adult-relationship pattern which is more and more being recognized as a motivational prerequisite of any kind of adult-controlled learning." The lower-class children in Milner's study perceived adults as predominantly hostile. The second advantage is an extensive opportunity to interact verbally with adults in the family. The latter point is illustrated by parental attitudes toward mealtime conversation, lower-class parents tending to inhibit and discourage such conversation, while middle-class parents encourage it.

Most traditional studies on child-rearing practices have been designed in terms of a psychoanalytic orientation. There is need for more data pertaining to other types of hypotheses. Findings such as those of Milner on opportunities for verbalization and the resulting effects upon reading readiness represent a step in this direction. Another possible source of future data is the application of the intensive observational techniques of psychological ecology developed by Barker and Wright (7) to widely diverse socioeconomic groups.

A sixth major approach involves research on the previously cited somatopsychological relationships (6). To date, little direct information is available on the precise operation of this class of factors in psychological development. The multiplicity of ways in which physical traits—whether hereditary or environmental in origin—may influence behavior thus offers a relatively unexplored field for future study.

The seventh and final approach to be considered represents an adaptation of traditional twin studies. From the standpoint of the question "How?" there is need for closer coordination between the usual data on twin resemblance and observations of the family interactions of twins. Available data already suggest, for example, that closeness of contact and extent of environmental similarity are greater in the case of monozygotic than in the case of dizygotic twins (cf. 2). Information on the social reactions of twins toward each other and the specialization of roles is likewise of interest (2). Especially useful would be longitudinal studies of twins, beginning in early infancy and following the subjects through school age. The operation of differential environmental pressures, the development of specialized roles, and other environmental influences could thus be more clearly identified and correlated with intellectual and personality changes in the growing twins.

Parenthetically, I should like to add a remark about the traditional applications of the twin method, in which persons in different degrees of hereditary and environmental relationships to each other are simply compared for behavioral similarity. In these studies, attention has been focused principally upon the amount of resemblance of monozygotic as contrasted to dizygotic twins. Yet such a comparison is particularly difficult to interpret because of the many subtle differences in the environmental situations of the two types of twins. A more fruitful comparison would seem to be that between dizygotic twins and siblings, for whom the hereditary similarity is known to be the same. In Kallmann's monumental research on psychiatric disorders among twins (20), for example, one of the most convincing bits of evidence for the operation of hereditary factors in schizophrenia is the fact that the degrees of concordance for dizygotic twins and for siblings were practically identical. In contrast, it will be recalled that in intelligence test scores dizygotic twins resemble each other much more closely than do siblings—a finding which reveals the influence of environmental factors in intellectual development.

Summary

The heredity-environment problem is still very much alive. Its viability is assured by the gradual replacement of the questions, "Which one?" and "How much?" by the more basic and appropriate question, "How?" Hereditary influences—as well as environmental factors of an organic nature—vary along a "continuum of indirectness." The more indirect their connection with behavior, the wider will be the range of variation of possible outcomes. One extreme of the continuum of indirectness may be illustrated by brain damage leading to mental deficiency; the other extreme, by physical characteristics associated with social stereotypes. Examples of factors falling at intermediate points include deafness, physical diseases, and motor disorders. Those environmental factors which act directly upon behavior can be ordered along a continuum of breadth or permanence of effect, as exemplified by social class membership, amount of formal schooling, language handicap, and familiarity with specific test items.

Several current lines of research offer promising techniques for exploring the *modus operandi* of hereditary and environmental factors. Outstanding among them are investigations of: (*a*) hereditary conditions which underlie behavioral differences between selectively bred groups of animals; (*b*) relations between physiological variables and individual differences in behavior, especially in the case of pathological deviations; (*c*) role of prenatal physiological factors in behavior development; (*d*) influence of early experience upon eventual behavioral characteristics; (*e*) cultural differences in child-rearing practices in relation to intellectual and emotional development; (*f*) mechanisms of somatopsychological relationships; and (*g*) psychological development of twins from infancy to maturity, together with observations of their social environment. Such approaches are extremely varied with regard to subjects employed, nature of psychological functions studied, and specific experimental procedures followed. But it is just such heterogeneity of methodology that is demanded by the wide diversity of ways in which hereditary and environmental factors interact in behavior development.

REFERENCES

1. ANASTASI, ANNE. *Psychological testing.* New York: Macmillan, 1954.
2. ANASTASI, ANNE. *Differential psychology.* (3rd ed.) New York: Macmillan, 1958.
3. ANASTASI, ANNE, & CORDOVA, F. A. Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. *J. educ. Psychol.*, 1953, **44**, 1–19.
4. ANASTASI, ANNE, & FOLEY, J. P., JR. A proposed reorientation in the heredity-environment controversy. *Psychol. Rev.*, 1948, **55**, 239–249.
5. ARSENIAN, S. Bilingualism in the post-war world. *Psychol. Bull.*, 1945, **42**, 65–86.
6. BARKER, R. G., WRIGHT, BEATRICE A., MYERSON, L., & GONICK, MOLLIE R. Adjustment to physical handicap and illness: A survey of the social psychology of physique and disability. *Soc. Sci. Res. Coun. Bull.*, 1953, No. 55 (Rev.).
7. BARKER, R. G., & WRIGHT, H. F. *Midwest and its children: The psychological ecology of an American town.* Evanston, Ill.: Row, Peterson, 1955.
8. BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, **51**, 239–263.
9. BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, **53**, 210–226.
10. DARCY, NATALIE T. A review of the literature on the effects of bilingualism upon the measurement of intelligence. *J. genet. Psychol.*, 1953, **82**, 21–57.
11. DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. *Amer. sociol. Rev.*, 1946, **11**, 698–710.
12. DOBZHANSKY, T. The genetic nature of differences among men. In S. Persons (Ed.), *Evolutionary thought in America.* New Haven: Yale Univer. Press, 1950. Pp. 86–155.
13. DOBZHANSKY, T. Heredity, environment, and evolution. *Science*, 1950, **111**, 161–166.
14. FORGUS, R. H. The effect of early perceptual learning on the behavioral organization of adult rats. *J. comp. physiol. Psychol.*, 1954, **47**, 331–336.
15. HALDANE, J. B. S. *Heredity and politics.* New York: Norton, 1938.
16. HARRELL, RUTH F., WOODYARD, ELLA, & GATES, A. I. *The effect of mothers' diets on the intelligence of the offspring.* New York: Bur. Publ., Teach. Coll., Columbia Univer., 1955.
17. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
18. HIRSCH, J., & TRYON, R. C. Mass screening and reliable individual measurement in the experimental behavior genetics of lower organisms. *Psychol. Bull.*, 1956, **53**, 402–410.
19. HORWITT, M. K. Fact and artifact in the biology of schizophrenia. *Science*, 1956, **124**, 429–430.
20. KALLMANN, F. J. *Heredity in health and mental disorder; Principles of psychiatric genetics in the light of comparative twin studies.* New York: Norton, 1953.
21. KING, J. A., & GURNEY, NANCY L. Effect of early social experience on adult aggressive behavior in C57BL10 mice. *J. comp. physiol. Psychol.*, 1954, **47**, 326–330.
22. LOEVINGER, JANE. On the proportional contributions of differences in nature and in nurture to differences in intelligence. *Psychol. Bull.*, 1943, **40**, 725–756.
23. LORENZ, K. Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen. *J. Orn., Lpz.*, 1935, **83**, 137–213; 289–413.
24. LUCHINS, A. S., & FORGUS, R. H. The effect of differential postweaning environment on the rigidity of an animal's behavior. *J. genet. Psychol.*, 1955, **86**, 51–58.
25. MELZACK, R. The genesis of emotional behavior: An experimental study of the dog. *J. comp. physiol. Psychol.*, 1954, **47**, 166–168.
26. MILNER, ESTHER A. A study of the relationships between reading readiness in grade one school children and patterns of parent-child interaction. *Child Develpm.*, 1951, 22, 95–112.
27. PASAMANICK, B., KNOBLOCH, HILDA, & LILIENFELD, A. M. Socioeconomic status and some precursors of neuropsychiatric disorder. *Amer. J. Orthopsychiat.*, 1956, **26**, 594–601.
28. SCHWESINGER, GLADYS C. *Heredity and environment.* New York: Macmillan, 1933.

29. SCOTT, J. P., & CHARLES, MARGARET S. Some problems of heredity and social behavior. *J. gen. Psychol.*, 1953, **48**, 209–230.

30. SCOTT, J. P., & FULLER, J. L. Research on genetics and social behavior at the Roscoe B. Jackson Memorial Laboratory, 1946–1951—A progress report. *J. Hered.*, 1951, **42**, 191–197.

31. SEARLE, L. V. The organization of hereditary maze-brightness and maze-dullness. *Genet. Psychol. Monogr.*, 1949, **39**, 279–325.

32. SNYDER, L. H. The genetic approach to human individuality. *Sci. Mon., N. Y.*, 1949, **68**, 165–171.

33. SNYDER, L. H., & DAVID, P. R. *The principles of heredity.* (5th ed.) Boston: Heath, 1957.

34. SPOERL, DOROTHY T. Bilinguality and emotional adjustment. *J. abnorm. soc. Psychol.*, 1943, **38**, 37–57.

35. THOMPSON, W. R., & MELZACK, R. Early environment. *Sci. Amer.*, 1956, **194** (1), 38–42.

36. TRYON, R. C. Genetic differences in maze-learning ability in rats. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39**, Part I, 111–119.

37. WHITING, J. W. M., et al. *Field guide for a study of socialization in five societies.* Cambridge, Mass.: Harvard Univer., 1954 (mimeo.).

38. WHITING, J. W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study.* New Haven: Yale Univer. Press, 1953.

39. WILLIAMS, JUDITH R., & SCOTT, R. B. Growth and development of Negro infants: IV. Motor development and its relationship to child rearing practices in two groups of Negro infants. *Child Develpm.*, 1953, **24**, 103–121.

40. WOODWORTH, R. S. Heredity and environment: A critical survey of recently published material on twins and foster children. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 47.

(Received September 18, 1957)

Psychological Review
Vol. 65, No. 4, 1958

HOW ARE INTERTRIAL "AVOIDANCE" RESPONSES REINFORCED?

O. H. MOWRER AND J. D. KEEHN[1]

University of Illinois

Beginning with a rather incidental observation reported by May (14) in 1948, a number of investigators (6, 22, 25) have shown that in a conditioning set-up, involving a noxious unconditioned stimulus, the incidence of intertrial (spontaneous) responses will increase markedly if such responses are allowed to function instrumentally, i.e., to "avoid," in the sense of postponing, the next "trial." This effect is not obtainable if the unconditioned stimulus is preceded by a warning signal (27), but does occur regardless of whether the instrumental intertrial response is the same as or different from the response produced by the unconditioned stimulus (13; cf. 18). The increased incidence of intertrial responses under the conditions described is in contrast to the previously reported observation that intertrial responses occur progressively *less* frequently in a conditioning situation wherein they cannot function instrumentally, i.e., where the intertrial interval is fixed or at least variable only by the experimenter (7, 19).

The discovery that an instrumental procedure of the kind described will increase the incidence of intertrial responses followed logically enough from the finding of Hunter (11), Brogden, Lipman, and Culler (3), and others that a response made to a warning (conditioned) stimulus is more readily and more reliably fixated if, when such a response occurs, the unconditioned stimulus is omitted, rather than being paired willynilly (as in "classical" conditioning) with the conditioned stimulus. And from this and related work it was soon evident that so-called avoidance learning is not a matter of simple (stimulus-substitution) conditioning, but that it involves two distinct stages or "levels": when a signal is paired with some painful stimulus, what an organism learns first is to *be afraid;* and then it learns, perhaps by a quite different process, what to *do* about its fear (16, 17). Yet, with respect to the learning of instrumental (avoidance) intertrial responses, there has been a tendency to adhere to the older one-stage stimulus-substitution type of interpretation. The purpose of the present paper is to show, on both empirical and logical grounds, how much more satisfactory is a two-factor type of explanation.

PREVIOUS RESEARCH AND THEORY

In the winter of 1946, Hart Westbrook made what seems to have been the first attempt to set up an experimental situation in which intertrial responses might function instrumentally in the manner already indicated. However, since negative results were obtained (probably because training was not continued long enough; see later section on Interval-Avoidance Learning), the details of this study need not be described here. Then, as already indicated, two years later May reported the occurrence of intertrial response learning in a situation designed to investigate quite a different problem. This decidedly incidental, but striking, finding was reported thus:

> In the above experiments the swinging door over the barrier [separating the two sides of

---

[1] On leave of absence, 1956–57, by benefit of a Ford Fellowship, from the Department of Psychology, American University of Beirut.

> a shuttle box] was never locked. The rats could cross and recross at will. An indefinite number of crossings was possible during the training and testing trials. Sixty to 90 sec. were allowed between the last spontaneous crossing and the next training or testing trial. Thus, on the training trials an animal could avoid shock for a considerable time by crossing and recrossing at intervals of less than 60 sec. Ten of the 12 experimental animals (Group A—door closed) and 11 of the controls learned this form of adaptive behavior and used it fairly regularly both on the training and on the test trials. Very few of the animals in B-Group (door open), on the other hand, learned it (14, p. 71).

In the summary of the study cited, May does not allude to this finding and elsewhere in his paper makes no attempt to interpret or analyze it. However, in a study by Bugelski and Coyer (6), briefly reported in 1950, the problem was attacked quite directly and with a theoretical issue in mind, implied by the title, "Temporal Conditioning vs. Anxiety Reduction in Avoidance Learning." A few years later, Bugelski described the procedure employed thus:

> Every 15 seconds a shock was applied in sufficient strength to force a rat over a hurdle. With repeated trials, the animals came to give almost perfect performance, jumping the hurdle in the general interval of 10 to 15 seconds, thus avoiding shock (5, p. 68).

In accompanying graphs Bugelski shows that, although rats learned to "shuttle" more readily under the conditions indicated (i.e., when the shock-shock interval was 15 seconds), another group of animals also solved the problem with a shock-shock interval of 60 seconds. Say Bugelski and Coyer of their study:

> The results support Hull's account of avoidance behavior (*Principles*) as antedating responses developing from escape reactions. Anxiety reduction and expectancy explanations need not be invoked for the responses studied (6, p. 265).

By this, these writers apparently mean that shuttling, when it occurs between trials, is just the conditioned (antedating) form of the escape reaction elicited by the shock which now, however, is produced by the experimental situation alone (without shock). Such an interpretation is, in a sense, classically Pavlovian and, as these writers point out, does not require that fear be posited as an intervening variable or that fear reduction be a form of reinforcement. Reasons for doubting the adequacy of this view will be given presently.

Then, in 1953, Sidman, using a different type of response and training procedure, also reported positive results:

> White rats were the experimental organisms, with lever pressing selected as the avoidance response. Shocks of a fixed 0.2-sec. duration were given to the animal through a grid floor at regular [commonly 5-sec.] intervals unless the lever was depressed. Each lever depression reset the timer controlling the shock, thus delaying its appearance [by 20 seconds or so]. . . . Only the initial downward press on the lever reset the timer; holding the lever down had no effect upon the occurrence of the shock.
>
> With no other contingencies between avoidance behavior and exteroceptive stimulation involved, approximately 50 animals have been successfully conditioned. . . . A striking characteristic of the initial curves is the abruptness with which the rate increases from its initial near-zero level. . . . With continued training the rate remains relatively stable not only within but also between sessions. Rates as high as 17 responses/min have been maintained by some animals during sessions totaling over 24 hours, with variations no greater than 0.1 responses/min appearing between the average rate for each session (25, pp. 157–158).

Sidman was inclined to interpret this striking effect as follows:

> The behavior generated by this procedure can be explained by a model which holds that avoidance responding increases in rate at the expense of other behavior that is depressed by shock. An equivalent statement, in reinforcement terms, is that the avoidance response is strengthened when it terminates incompatible behavior that has been paired with shock (8, 23) [i.e., interval responses, other than the "correct" one, which do not postpone shock, get punished and thus tend to arouse fear]. Several lines of evidence indicate that the

avoidance rate is not simply some form of temporal conditioning in which the responses are triggered-off by the passage of a time interval (25, p. 158).

Because of the brevity of the paper just cited, the precise nature of Sidman's theoretical position was not entirely clear, so we turn to the paper by Schoenfeld which Sidman cites. This paper is long and intricate and cannot be summarized here; but the following sentences, from the summary, are perhaps the most relevant ones for present purposes:

> Following several objections raised against [the] anxiety-reduction hypothesis, an alternative explanation of avoidance is offered in which the response is conceived to be always a stimulus-terminating one. . . . It is argued that the avoidance response terminates [aversive] stimulus compounds in which proprioceptive and tactile stimuli are important components. . . . If this formulation is correct, avoidance conditioning proves to be a form of escape training, and its avoidance function is incidental to its stimulus-termination function (23, p. 97).

That the Schoenfeld position is, however, open to varying interpretations is indicated by the following comment by Kamin:

> The usual assumption has been that spontaneous responses result from conditioning to generalized apparatus cues (cf. 9, p. 76; 23, p. 89) (12, p. 71).[2]

While the theoretical picture was thus somewhat confused as of 1954, it is fair to say that the prevailing tendency was to explain the reinforcement of instrumental intertrial responses as *not* involving the reduction of a temporally conditioned fear (or anxiety), and perhaps not involving the reduction of fear at all.

EXPERIMENTAL DISPROOF OF CLASSICAL CONDITIONING INTERPRETATION

The fear-reduction hypothesis has been so successful in accounting for avoidance learning in general, and in resolving paradoxes generated by other theories, that there is strong presumption that it can also explain the learning of instrumental intertrial responses. But at once there is at least a superficial difficulty. In "ordinary" avoidance learning, i.e., where there is a definite and specific warning signal, it is simple and reasonable to suppose (*a*) that fear becomes conditioned to this signal and (*b*) that any response which turns the signal off (and averts the unconditioned stimulus) will be reinforced by fear reduction. But, in the case of instrumental intertrial responses, there is no such specific warning signal, so that the fear-reduction hypothesis, if applicable at all, must apply in some rather subtle and special way. In later sections, this inference will be considered in detail; but first a simple experiment will be reported which clearly shows the inadequacy of the classical conditioning type of explanation.

Kamin has already been quoted as saying: "The usual assumption has been that spontaneous responses result from conditioning to generalized apparatus cues." If the classical conception of conditioning were valid, then, every time a subject experienced electric shock in a given experimental situation and made a specific response thereto, there would be an increased tendency for the subject to make that response to the stimulus compound which immediately precedes the shock. Since it is the experimental situation as a whole, rather than a more specific stimulus, which here precedes shock and thus becomes, in effect, the conditioned stimulus, the subject should soon start making the response in question to that situation "spontaneously," i.e., without a specific cue or signal.

---

[2] Solomon and Brush (33, p. 266) allude to a study of intertrial avoidance learning reported (as a personal communication) by Milner in 1955. In only one respect (which will be mentioned later) does this study extend what had been learned from earlier experiments. Solomon and Brush also give a comprehensive review of research on avoidance learning where the CS-US interval is systematically varied (with different subjects). However, this line of work, while germane, is not directly relevant to the present discussion and will not be specifically cited.

On the other hand, if one accepts the two-factor version of avoidance learning, one must assume that, in order for intertrial responses to become more frequent, they must be reinforced by the occurrence of fear reduction. In the one case, mere contiguity of stimulation (situation + shock) would provide the conditions necessary for the increased occurrence of intertrial responses; whereas, in the other case, something more would be required: these responses would have to be in some sense functional, instrumental, problem solving. In the experiment to be described shortly, these two hypotheses are put to the test; but, before describing that experiment, it will be necessary to say a word about apparatus and procedure.

In the instrumental intertrial learning experiments already alluded to, Westbrook and Schaefer used a wheel-turning response, May and Bugelski and Coyer used shuttling, and Sidman used bar pressing. In casting about for an avoidance response that might be more "natural" for rats than either bar pressing or wheel turning and also free of the conflict that necessarily exists (at least in the beginning) in a shuttling set-up, one of the present writers (J. D. K.) tried out a revolving cage similar to that employed in the experiment of Brogden, Lipman, and Culler (3) and found it highly satisfactory. With the response now merely a short *run* (which revolves the cage but otherwise gets the subject nowhere), rats show excellent interval-avoidance learning; and with a few refinements the apparatus has proved itself to be reliable, flexible, and well adapted to the investigation of a wide range of related problems. The refinements are these: (*a*) a simple, silent "lock" which allows the cage to move easily in one direction but not in the other; (*b*) a device which quietly resets an interval timer whenever the cage is turned (in the "correct" direction) by a minimum of two or three inches; and (*c*) articulation of the revolving cage with a cumulative graphic recorder (described elsewhere, 15). Thus, most of the procedure and all the recording are made automatic and put beyond the influence of experimenter produced variables.

Figure 1 shows the performance typical of an albino rat (a 140-day-old male) in the apparatus here described, under the following conditions. For one hour on three successive days this rat had been put into the revolvable cage, with no shock, and allowed to habituate. What little spontaneous running occurred in the beginning soon disappeared, so that on the third day the rat, in the course of an hour, turned the cage less than two revolutions, as can be seen by the virtually horizontal "base-line" in the lower left-hand corner of Fig. 1. Then, after the three daily sessions of habituation, the rat was subjected to daily training sessions, again of one-hour duration each, under the following conditions: a light electric shock from the grillwork which constitutes the "floor" of the revolvable cage came on every 20 seconds and remained on until the rat ran and revolved the cage far enough (roughly two inches) to turn off the shock, *unless* such a run occurred *between* shocks, in which event the timing mechanism reset and started another 20-second cycle, without shock. As will be seen in Fig. 1, this subject, during the first experimental session, received 37 shocks (indicated by the lateral spurs on the cumulative response line). Since a total of 180 shocks (3 to the minute for 60 minutes) would have been received during the experimental

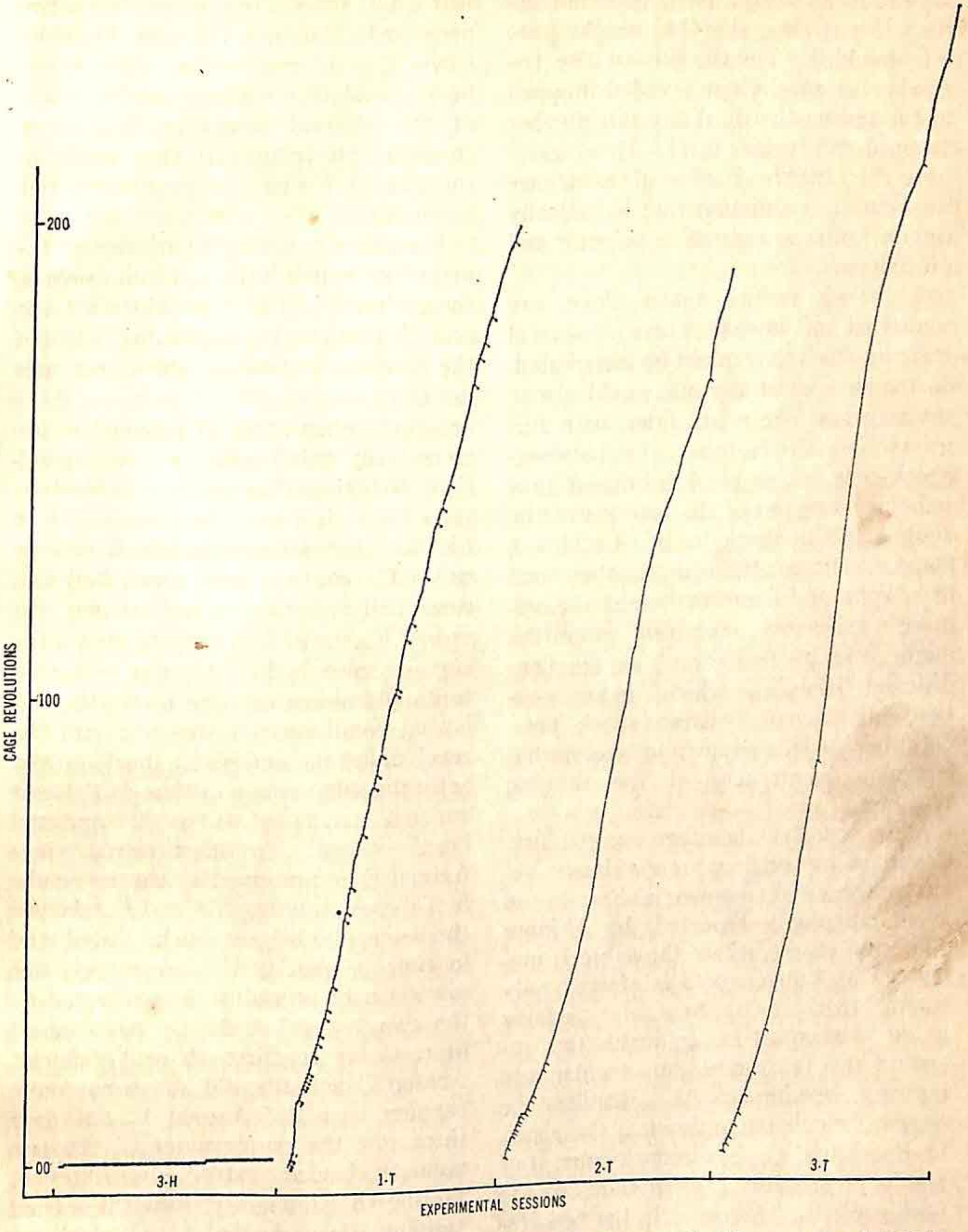


FIG. 1. Cumulative records of an albino rat's performance in a revolvable cage when between-trial ("spontaneous") runs postpone the next shock by 20 seconds. An activity baseline, representing the amount of running by the subject during the third of three daily habituation sessions of one hour each, under conditions of no shock, is shown in Graph 3-H. Graphs 1-T, 2-T, and 3-T represent the amount of running during hourly sessions on each of three successive "training" days. Shocks were applied only when the subject failed, for 20 seconds, to revolve the cage by the prescribed amount (two or three inches). Shocks are indicated by the lateral spurs on the underside of the graphs.

session if no interval running had occurred, it is clear that 143 shocks were thus avoided. On the second day the number of shocks not avoided dropped to 15, and on the third day this number dropped still further to 11. Here, manifestly, is highly successful avoidance learning, in a situation that is basically simple both as regards apparatus and procedure.

But such results, taken alone, are somewhat ambiguous. There are several ways in which they might be interpreted. In the absence of any other evidence to the contrary, one might infer, with Bugelski and Coyer, that the between-shock runs are simply conditioned (antedating) versions of the reaction of the subject to the shock itself. Lacking a more specific conditioned stimulus, such as a light or buzzer to herald the approach of shock, the *total situation* might here be thought of as the conditioned stimulus, which, being continuously present between shock presentations, elicits numerous, somewhat irregular occurrences of the running response.

Or, consider another possibility. There is an obvious resemblance between the curves shown in Fig. 1 and those commonly reported, by Skinner (31) and others, where the subject, motivated by hunger, gets reinforced only intermittently (with food) for pushing a bar but continues to make this response at a fairly rapid and regular rate between reinforcements. Unlike the hunger, which is continuous, the shock in this situation is discontinuous; but fear is presumably present much of the time while the subject is in the revolvable cage. So one might think of the response to shock-plus-fear as generalizing to fear alone and thus occurring many times on the basis of the sheer habit strength (or reflex reserve) generated solely by the reinforcement provided by escape from the shock-plus-fear on the occasions when shock was presented. Here, as in the Bugelski-Coyer type of explanation, there would be no need to posit any reinforcement of the interval responses themselves, through fear reduction; they could be thought of instead as pure extinction responses.

Fortunately, a simple variation of the procedure which leads to results such as those shown in Fig. 1 provides a fairly crucial means of determining whether the interval responses which are here exhibited are reinforced only on those occasions when shock is received *or* are secondarily reinforced, by fear reduction, each time they occur. Typical results for such a procedure are shown in Fig. 2. Here an animal, which may be called C (control) and which had had three daily sessions of habituation just as had E (the animal used to obtain the curves shown in Fig. 1), was subjected to daily training sessions under the following conditions: it was put into the revolvable cage and given shocks at exactly the same points in time as those at which E had failed to "avoid" and had been shocked. In other words, since Animal E had received 37 shocks on the first day of training, Animal C received the same number of shocks distributed in time in exactly the same way; the same sort of procedure was repeated on the two remaining days. As a result of receiving the first 10 or 12 shocks, Animal C actually did somewhat more running than did Animal E, but soon thereafter the performance of the two animals became strikingly different. During the course of Session 1-T (first training day), Animal C, although receiving the same number (and approximate duration) of shocks, turned the cage only 94 revolutions as opposed to 198 revolutions for Animal E; and, in Session 3-T, the discrepancy had become much wider still: only 10 revolutions for Animal C as opposed to 249

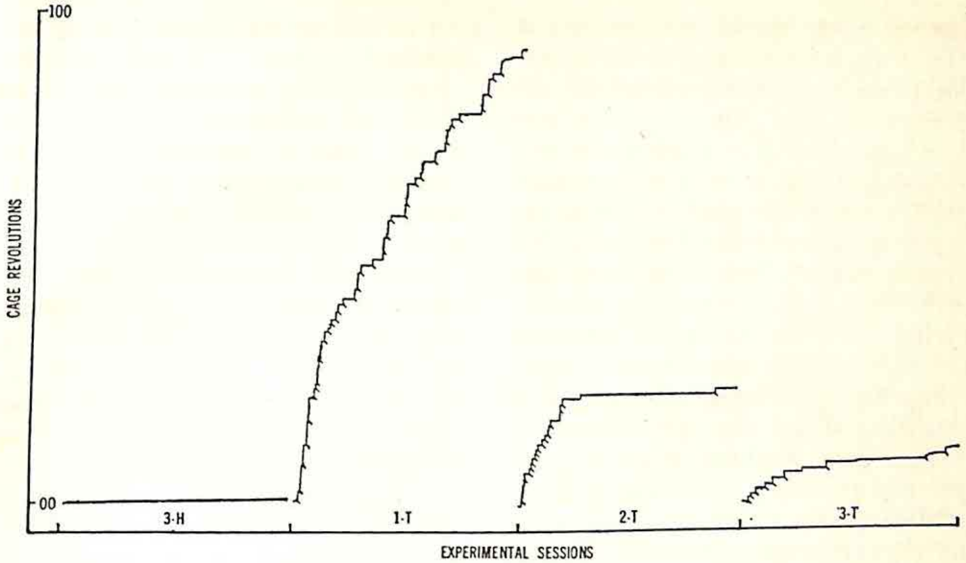


FIG. 2. Cumulative records of an albino rat's performance in a revolvable cage when a run (of two or three inches) turns off the shock (received from the cage floor) but, when occurring spontaneously, does not postpone the next shock. The four graphs are labeled as in Fig. 1. The shocks, equal in intensity and number, are distributed in time just as were those received by the subject whose performance is shown in Fig. 1. This type of procedure provides a control for certain hypotheses, cited in the text, which have been advanced to account for the type of interval-response avoidance learning shown in Fig. 1.

revolutions for Animal E. As Curve 3-T shows, when the shock came on, Animal C would still run far enough to turn the shock off (or perhaps a little further), but there was virtually no interval running. This animal had learned that the shocks were not avoidable, that they came regardless of what was done in the intervals between shocks; so, there being no gain from such behavior, the animal, sensibly enough, abandoned it. This outcome, substantiated by results obtained with three other pairs of animals, supports the hypothesis that interval responses, if they are to be perpetuated, must be reinforced in some manner other than by the reinforcement provided by the mere pairing of situation and shock.

INTERVAL AVOIDANCE LEARNING FURTHER ANALYZED AND INTERPRETED

From the empirical results just reported, it is evident that intertrial avoidance learning cannot be adequately explained in terms of classical conditioning. Some more intricate interpretation is obviously called for; the one which most readily suggests itself involves the now well authenticated concept of *stimulus trace* (1, 4). In an experimental situation of the kind under discussion, each response, whether forced (by shock) or spontaneous, presumably sets up a neural reverberation ("immediate memory") which decays to zero (or at least to asymptote) in roughly 30 seconds. If, therefore, electric shock be delivered to the subject whenever, let us say, the 20-second point on the stimulus trace is reached, one would expect that fear would become conditioned to this point on the trace and, in generalizing forward, would motivate the subject to make the response in question during the intertrial (20-second) interval, thus averting the impending shock. And, since each new occurrence of the response would, so to say, "reset" the trace, there would be a reduction in fear and hence a reinforcement of the response itself (cf. 7, 20).[3]

Reasonable as it is a priori, this hypothesis did not, at first, seem to have much empirical justification. The Westbrook study was undertaken with this hypothesis explicitly in mind; and when preliminary work failed to reveal the expected effect, i.e., showed no tendency for the intertrial response to become temporally conditioned, the study was discontinued and the hypothesis questioned. May, as already noted, was apparently the first to obtain good intertrial avoidance learning, but he made no attempt to account for it theoretically. Bugelski and Coyer, to be sure, reported (6) that, with repeated trials, the animals came to give almost perfect performance, jumping the hurdle in the general interval of 10 to 15 seconds, thus avoiding shock; but these writers advanced the now untenable antedating hypothesis (of Pavlov and Hull). Sidman, finding no evidence of temporal conditioning in his own early researches (see Fig. 3), adopted an explanation, derived from Schoenfeld, to the effect that the "correct" interval response is learned because other intertrial responses get punished (in the sense of failing to postpone shock) and therefore, relative to the "correct" response, become inhibited. Moreover, since the occurrence of the correct response stops or prevents the occurrence of other (fear-producing) responses, this response will have a tendency to reduce fear and thus be reinforced—an interpretation which makes no use whatever of the stimulus-trace concept. (Sidman did not, to be sure, use precisely the above terminology; but, in light of considerations to be advanced in the next section, this paraphrasing seems legitimate.)

Schaefer, in the recent study already cited, likewise failed to find evidence of temporal conditioning in his subjects—in fact, their intertrial responses tended decidedly to pile up (as in Fig. 3) toward the beginning of the intertrial interval rather than toward the end thereof. And on the basis of this finding, Schaefer advanced an explanation of intertrial avoidance learning which, while in effect something like the one put forward by Sidman, is at least diffently phrased. According to Schaefer's conjecture, what the subject in an experiment of this kind does is to *discriminate* between two conditions: situation with correct response being made and situation with this response *not* being made. And, since the rat is never shocked while thus responding and *is* shocked when responding differently or not at all, then the rat, feeling safe while responding correctly and fearful at all other times, is differentially reinforced on each occurrence of the correct response. In other words, by making a particular response, the animal, so to say, converts the situation from a dangerous one to a safe one and is thereby rewarded. Here the dilemma mentioned earlier—that there is no explicit danger signal in a situation of this kind whose termination can provide the specific occasion for fear reduction and reinforce-

---

[3] Imagine an experimental situation in which a relatively loud tone came on immediately after the subject had responded in a specified way and then gradually softened—or a patch of light which, upon occurrence of the response, became suddenly large and then gradually diminished in size. Suppose also that the subject never received shock when the stimulus was large (or loud) but was shocked when it diminished beyond a certain point. It would be in no way surprising if the subject, under these conditions, learned to use such a stimulus as the basis of successful avoidance behavior. Our conjecture is that, in interval-avoidance learning of the kind under discussion, something comparable is involved, the principal difference being that the changing cue, instead of being an external stimulus, is a stimulus *trace* produced by the last preceding instrumental (overt) response.

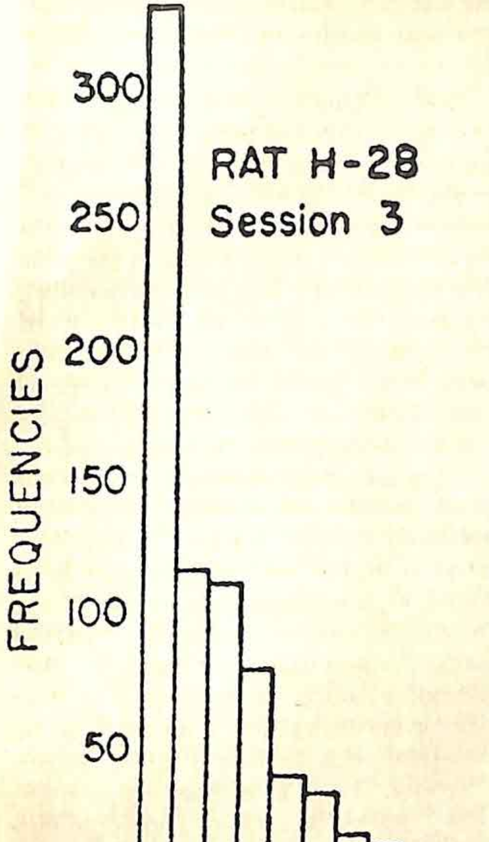


Fig. 3. Distribution of intertrial responses made by a rat during Session 3 of an experiment reported by Sidman (26). The abscissa represents seconds of time elapsing between the preceding response and the next one. When the subject delayed by as much as 20 seconds between responses, it then received an electric shock of brief (0.2 sec.) duration. As will be noted, most responses occurred shortly after the preceding one, rather than piling up in the true "danger zone," i.e., during the latter part of the intertrial period (see text).

ment of the avoidance response—is resolved by making the situation as a whole, *without correct responding,* the "danger signal," which is then "turned off" whenever and while the animal makes the requisite response.

This is an eminently satisfactory theory (and is stated more simply and plausibly than Sidman's somewhat equivalent one); but it cannot, alone, account for all the now known facts. In 1954, in a paper entitled "The Temporal Distribution of Avoidance Responses," Sidman reported that "after continued training" there is a tendency for a "well developed time-discrimination" to appear (26, p. 401). In two recent papers, Boren and Sidman (2, 29) show that, if training is continued long enough (to or beyond approximately 100 hours), then interval-avoidance responses do indeed take on the expected temporal distribution, i.e., they tend to pile up near the remote end of the intertrial interval, rather than toward the beginning thereof (see Solomon and Brush's account of Milner's work, 33, p. 266). Hence, even though the Sidman-Schaefer type of explanation be valid for the results first obtained in interval-response avoidance learning, such an explanation must obviously be supplemented if one is also to account for the opposite type of temporal distribution of responses obtained after more extended training. One possibility, therefore, is that, in experimentation of the kind under discussion, the *first* discrimination that is developed is a comparatively gross one, i.e, between experimental situations with and without the correct response being in progress, and that a *second* stage of discrimination later develops in which the subject discovers that situation-without-correct-responding is dangerous *only* after an interval of time has elapsed.

There is, however, also the possibility that the stimulus-trace type of explanation can account for *all* the facts. Let us assume that fear is most strongly conditioned to a point on the stimulus trace 20 seconds removed from the event that sets up the trace (i.e., the preceding correct response, either forced or spontaneous). Let us further assume, as we reasonably may, that fear (at least in the beginning) *generalizes far forward* on the trace. If, then, as already posited, occurrence of the correct response during the intertrial period sets up a new trace and reduces fear, one might expect this response to occur just as soon as the subject *starts* being afraid. In other words, there is nothing in the nature of the situation which would require that the subject wait until *maximally* afraid before reacting defensively. Therefore, the temporal distribution of the incidence of the correct overt response would not be expected necessarily to provide a faithful picture of the temporal pattern of fear intensity. Occurrence of the correct overt intertrial response might very well pile up, not where the fear would be maximally intense, but rather where it is first experienced.

Eventually, of course, one would expect discrimination to develop and to offset this original overgeneralization, so that fear would be first experienced during the intertrial period, not far forward (due to generalization along the trace), but instead toward the latter part of the period where the objective danger really is.

In a study published in 1956 under the title "Time Discrimination and Behavioral Interaction in a Free Operant Situation," Sidman has shown that it is a simple matter to get rats to respond with considerable accuracy to a 20-second temporal interval when they are thirst-motivated and water-reinforced for pressing one bar, at the appropriate interval, after having pressed (on cue) another bar. Thus, as Sidman observes:

> The procedure is an operant analogue of Pavlovian trace conditioning, with the qualification that the "trace" is initiated not by an exteroceptive stimulus but by the organism's own behavior. . . . The response probability increases with the passage of time, not because of the dissipation of inhibition [as Hull (10) is credited with holding], but rather because the situation becomes more like that prevailing at the time of reinforcement (28, p. 469).

At first it may appear strange that a free-operant type of behavior (bar pressing) can be temporally conditioned so readily when the subjects are thirst-motivated; whereas, as already noted, in an avoidance-learning situation the same free operant for a long time occurs much too soon, i.e., long before the time when the unconditioned stimulus will actually occur. But there is an instructive difference: in the situation when the subjects are thirst-motivated, responses that occur too soon go unrewarded, i.e., do not produce water and are thus extinguished relatively rapidly; whereas, in the situation when the subjects are *fear*-motivated, each and every occurrence of the "correct" response during the intertrial period, no matter how early in that period, produces a decrement in fear (however slight) and thus tends to be self-reinforcing, rather than self-extinguishing. Here, as in many other situations, we see the importance of differentiating between primary and secondary drive and drive reduction.

TWO DIFFERENT CONCEPTIONS OF "AVERSIVE" STIMULATION

In earlier sections of this paper, we have seen how numerous and determined the attempts have been to account for various forms of avoidance behavior without recourse to the concept of fear as an intervening variable. There is, of course, every justification in science for trying to push the principle of parsimony as far as possible. But it must be remembered that sometimes a given type of parsimony proves to be quite impossible and has to be replaced by a more complex conceptual scheme. It is the present writers' belief that a simple one-step (S–R) conception of behavior is inadequate and that, at the very least, we must posit a two-step S–r:s–R model, where S and R, as in the simpler S-R scheme, are "objective" stimulation and response, but where, in addition, r and s are intervening variables, of which *fear* (as both a reaction and as a drive) is a prime example.

The attempt of Sheffield to interpret the Brogden-Lipman-Culler experiment (previously cited) solely in terms of the contiguity (stimulus-substitution) principle and thus to eliminate the necessity of positing fear as an intervening variable illustrates the difficulties which such efforts commonly encounter. Says Sheffield:

> The findings for avoidable shock offer no evidence for a strengthening effect of avoidance and follow the conventional result in Pavlovian experiments. Reinforcement by shock strengthened conditioned running; omission of shock led to extinction (24, pp. 174–175).

The point can be illustrated by referring back to Fig. 1. There it will be seen that, in Sessions 2-T and 3-T, in each of the two longest runs of interval-responding, there is, indeed, a tendency for this behavior to extinguish (as evidenced by the flattening of the cumulative response curve) and to have to be eventually reinforced by shock presentation. Here, it might seem, is clear evidence that shock avoidance is not reinforcing and that the response of running is indeed dependent upon the occasional occurrence of the shock. The difficulty, of course, is the one that always arises in such a situation as a result of speaking as if only *one* response were involved. The present writers assume that shock is, of course, occasionally necessary to reinforce the *fear* response, but that the *running* response is reinforced by fear reduction not by shock. It is our assumption that, whenever enough fear is present to motivate intertrial running, there will be an ensuing experience of fear reduction, with the result that this behavior is always reinforced. It is, rather, the underlying fear that extinguishes and has to be eventually reinforced by presentation of shock. Our assumption is that, if fear could be maintained at an appropriate intensity by some other means, the intertrial running would *never* extinguish. This assumption is supported by the recently reported finding of Sidman, Herrnstein, and Conrad that:

> The effect of occasional free shocks, delivered independently of the monkeys' behavior, was to increase the rate of avoidance responding. . . . The effect of the free shock persisted over as long as 300 experimental hours during which time no other shocks were delivered (30, p. 557).

The authors' interpretation of this effect, in terms of Skinner's theory of superstitious behavior, seems to us contrived and nonparsimonious.

In an unpublished paper, Sidman, Herrnstein, and Boren try—but, we believe, unsuccessfully—to develop a conceptual scheme which will account for all forms of avoidance learning without recourse to the concept of fear as a motivating and reinforcing agent. In essence, the position is pristine behaviorism and stems directly, as the authors acknowledge, from the empirical researches and methodological stance of B. F. Skinner. In his 1953 book, *Science and Human Behavior,* Skinner states his position in this connection with stark simplicity. The chapter on "Emotion" begins with this sentence: "The 'emotions' are excellent examples of the fictional causes to which we commonly attribute behavior" (32, p. 160). Then the familiar, but highly debatable, assertion by William James is approvingly quoted to the effect

> that we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble because we are sorry, angry, or fearful, as the case may be (32, p. 160).

And, later in the chapter, there is a section specifically headed "Emotions Are Not Causes."

Rarely are theoretical issues (in psychology) so beautifully explicit and "clean." The present writers, along with many others, take the position that the emotion of fear *is* causal; that a one-step (simple S–R) psychology will not work (17, 21); that, as a minimum, two causal steps are necessary to account for behavior; and that fear, in the case of so-called avoidance behavior, is an essential intermediate cause or determinant.

On the other hand, Skinner and those of like persuasion hold that this is not at all the case and that a "chaining interpretation," not greatly different from the earlier Watsonian notion of reflex chaining, is "capable of handling most of the data and of extension to other findings, while at the same time maintaining a degree of parsimony and internal consistency that we were unable to achieve by means of the alternative considerations" (unpublished manuscript by Sidman-Herrnstein-Boren).

This is not the time or place to enter into a detailed critique of the chaining interpretation, but one particularly questionable aspect thereof may be pertinently noted, namely, the tendency to speak of stimuli which have been paired with punishment as being aversive. This, we believe, involves a logically indefensible attempt to avoid the more likely possibility that such stimuli, after conditioning, become aversive because, and only in the sense that, they now elicit *fear* and that it is this latter state, and not the eliciting stimulus, which provides the aversive, motivating element in the situation (cf. Solomon and Brush, 33, p. 245 ff.).

SUMMARY

It is now empirically established (*a*) that responses which occur between trials in a noxious conditioning situation occur progressively less frequently if trials occur ineluctably, i.e., at fixed intervals, and (*b*) that between-trial responses occur progressively *more* frequently when the next trial is delayed each time such a response is made. There is, of course, no mystery as to why intertrial (spontaneous) responses disappear in the former situation: they accomplish nothing and, if they happen to occur late in the intertrial interval, the ensuing noxious stimulus acts, in effect, as a punishment.

But there has been considerably less agreement as to how it is, under the second set of conditions mentioned, that intertrial responses are reinforced and perpetuated. A number of writers have taken the position that here the intertrial response (made to the experimental situation as a whole, rather than to a specific danger signal) is just the antedating (classically conditioned) form of the response made to the noxious, unconditioned stimulus on the trial. In the present paper various considerations are cited which indicate that this explanation is too simple and that a more complex explanation, involving the intervening advents of fear arousal and fear reduction, is necessary and generally sufficient.

Intertrial avoidance learning, as a laboratory phenomenon, deserves continued study, both because of its intrinsic theoretical interest and because of its rather patent relationship to compulsive (neurotically repetitive) behavior and to the psychology of *work*.

REFERENCES

1. ANDERSON, O. D., & LIDDELL, H. S. Observations on experimental neurosis in sheep. *Arch. Neurol. Psychiat.*, 1935, 34, 330–354.
2. BOREN, J. J., & SIDMAN, M. Maintenance of avoidance behavior with intermittent shocks. *Canad. J. Psychol.*, 1957, 11, 185–192.

3. BROGDEN, W. J., LIPMAN, E. A., & CULLER, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.,* 1938, **51**, 109–117.
4. BROWN, J. S. A note on a temporal gradient of reinforcement. *J. exp. Psychol.,* 1939, **25**, 221–227.
5. BUGELSKI, B. R. *The psychology of learning.* New York: Holt, 1956.
6. BUGELSKI, B. R., & COYER, R. A. Temporal conditioning vs. anxiety reduction in avoidance learning. *Amer. Psychologist,* 1950, **5**, 264–265 (Abstract).
7. COPPOCK, H., & MOWRER, O. H. "Spontaneous" responses as "rehearsal": A study of "overt thinking" in animals. *Amer. J. Psychol.,* 1947, **60**, 608–616.
8. HEFFERLINE, R. F. An experimental study of avoidance. *Genet. Psychol. Monogr.,* 1950, **42**, 231–334.
9. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
10. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
11. HUNTER, W. S. Conditioning and extinction in the rat. *Brit. J. Psychol.,* 1935, **26**, 135–148.
12. KAMIN, L. J. Traumatic avoidance learning: The effects of CS-US interval with a trace-conditioning procedure. *J. comp. physiol. Psychol.,* 1954, **47**, 65–72.
13. KEEHN, J. D. On the nonclassical nature of avoidance behavior. *Amer. J. Psychol.,* in press.
14. MAY, M. A. Experimentally acquired drives. *J. exp. Psychol.,* 1948, **38**, 66–77.
15. MOWRER, O. H. A cumulative graphic work-recorder. *J. exp. Psychol.,* 1943, **33**, 159–163.
16. MOWRER, O. H. On the dual nature of learning—a re-interpretation of "conditioning" and "problem solving." *Harvard educ. Rev.,* 1947, **17**, 102–148.
17. MOWRER, O. H. Two-factor learning theory reconsidered, with special reference to secondary reinforcement and the concept of habit. *Psychol. Rev.,* 1956, **63**, 114–128.
18. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.,* 1946, **39**, 29–50.
19. MOWRER, O. H., & LAMOREAUX, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.,* 1951, **58**, 196–212.
20. MOWRER, O. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. Soc. Psychol.,* 1948, **43**, 193–200.
21. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
22. SCHAEFER, V. H. Avoidance conditioning in the absence of external stimulation: Some experimental and genetic parameters. Unpublished doctoral dissertation, Univer. of Illinois, 1957.
23. SCHOENFELD, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. H. Hock & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950.
24. SHEFFIELD, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.,* 1948, **41**, 165–177.
25. SIDMAN, M. Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science,* 1953, **118**, 157–158.
26. SIDMAN, M. The temporal distribution of avoidance responses. *J. comp. physiol. Psychol.,* 1954, **47**, 399–402.
27. SIDMAN, M. Some properties of the warning stimulus in avoidance behavior. *J. comp. physiol. Psychol.,* 1955, **48**, 444–450.
28. SIDMAN, M. Time discrimination and behavioral interaction in a free operant situation. *J. comp. physiol. Psychol.,* 1956, **49**, 469–473.
29. SIDMAN, M., & BOREN, J. J. Avoidance behavior maintained by shock-correlated variable response-shock intervals. *J. comp. physiol. Psychol.,* in press.
30. SIDMAN, M., HERRNSTEIN, R. J., & CONRAD, D. G. Maintenance of avoidance behavior by unavoidable shocks. *J. comp. physiol. Psychol.,* 1957, **50**, 553–557.
31. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century Crofts, 1938.
32. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.
33. SOLOMON, R. L., & BRUSH, ELINOR S. Experimentally derived conceptions of anxiety and aversion. *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1956.

(Received September 20, 1957)

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated
Acc. No

Psychological Review
Vol. 65, No. 4, 1958

THE DERIVATION OF SUBJECTIVE SCALES FROM JUST NOTICEABLE DIFFERENCES[1]

R. DUNCAN LUCE

Department of Social Relations, Harvard University

AND WARD EDWARDS

Operator Laboratory, Air Force Personnel and Training Research Center

The study of cumulated jnd scales began with Fechner; Fechner's law is such a scale. Psychophysicists have been deriving such scales and comparing them with scales derived in other ways, notably by fractionation, ever since, and a lot of controversy has resulted. The controversy is particularly hot at present because Stevens and Galanter (**16**) and Stevens (**14**) have assembled a lot of data which indicate that cumulated jnd scales do not agree with magnitude scales derived by other methods for intensity continua such as loudness, brightness, and pain.

Unfortunately, Fechner's procedure for cumulating jnds, which has been widely defended but not widely applied since his day, rests on an assumption which is inconsistent with one of his definitions. This means that cumulated jnd scales developed by his procedure are incorrect, and so comparisons between them and other kinds of scales are meaningless.

This paper begins by showing that Fechner's method contains internal contradictions for all but a few special cases, and that it cannot be rescued by minor changes. It goes on to derive a new and mathematically appropriate method for cumulating jnd's. This method turns out to be the simplest possible one: you can best cumulate jnd's simply by adding them on top of each other, like a stack of plates. Unfortunately, the detailed mathematical equivalent of this very simple operation is often fairly complicated. A simple but sometimes tedious graphic procedure, however, is readily available—and indeed has customarily been used by most scientists when developing cumulated jnd scales. This paper ends by discussing practical applications of this method, the relation it bears to scaling methods based on the law of comparative judgment, and the current controversy about scaling methods in psychophysics.

The model of a sensation scale. The psychophysical model of a sensation scale is a mathematical model; a sensation scale is an intervening variable. The rules by which sensation scales should be constructed are to some degree arbitrary, limited by logic, convenience, intuition, and best fit to data.

---

[1] Work on this paper began when Luce was a Fellow at the Center for Advanced Study in the Behavioral Sciences. Later work was supported by: the Office of Naval Research, through contracts with Columbia University; the Behavioral Models Project, Bureau of Applied Social Research, Columbia University; and the Air Force Personnel and Training and Research Center under Project 7737, Task 27107. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States government. This paper will be listed as Publication A-243 of the Bureau of Applied Social Research, Columbia University.

We are grateful to J. K. Adams, E. G. Boring, L. S. Christie, E. H. Galanter, A. Hastorf, C. K. Kluckhohn, W. J. McGill, F. Mosteller, and S. S. Stevens for advice and help at various stages in the preparation of the paper.

The model of a sensation scale goes as follows. Corresponding to many of the major subjective dimensions of change of sensory experience, there are primary physical dimensions of change (e.g., pitch and frequency, loudness and amplitude, etc.). Once parametric conditions for significant variables have been specified, we assume that a single-valued, monotonic, everywhere differentiable (smooth) function exists that relates the subjective dimension to its corresponding physical dimension. From here on, we shall use the words "dimension" and "continuum" interchangeably; we shall usually talk about a stimulus continuum and its corresponding sensory continuum.

That is the model, and it is very easy to state. The big difficulty comes when we try to decide how to fit data to it. All methods for doing this must introduce definitions and assumptions beyond those listed in the previous paragraph. These differ from one method to another.

The oldest sensory scaling method, Fechner's, is based upon a further condition that says that any jnd on a given sensory continuum is subjectively equivalent to any other jnd on that continuum. Whether this added condition is to be interpreted as merely a definition of the scale under consideration or as an assumption is a matter of opinion. Textbooks usually say Fechner "assumed" that all jnd's for a sensory dimension are equal to one another (1). It is not easy to know what he had in mind, but judging by his writings he probably did view it as an assumption having implications beyond scale construction. Since it is not directly observable and since its indirect consequences are highly debatable, others since Fechner have suggested that it might better be viewed as a definition. It is our view that this is the more sensible position; it certainly is the one for us to take in this paper since our point is a logical one, not a substantive one. If equality of jnd's is taken as a definition, then it cannot be proved or disproved by any kind of empirical evidence. An experiment, for instance, that showed that a tone 20 jnd's loud is not half as loud (according to fractionation judgments) as a tone 40 jnd's loud would not have any relevance to what we may call Fechner's definition; it would only show that the kind of sensation scale implied by his definition does not agree with the kind implied by the definitions used in fractionation experiments. The issue, then, becomes what are the different scales useful for, and what is their relationship one to another. The latter part of this paper touches briefly on this problem.

The main purpose of the paper is to explore the consequences of Fechner's definition. Therefore, we must be certain of its meaning. It clearly does not mean, for instance, that all jnd's for loudness contain the same number of physical units. It is only on the sensory continuum, not on the stimulus continuum, that jnd's are defined as equal to one another. Furthermore, this definition holds only if all stimulus properties except those on the primary stimulus continuum remain constant. So there is no reason, for instance, to expect that the subjective size of a loudness jnd at 1,000 cycles per second (cps) should be the same as that of a loudness jnd at 4,000 cps. Of course, it would be pleasant if they were equal in size, but the model does not require it.

From here on we will talk about two kinds of jnd's. A sensation jnd is the magnitude of a jnd as measured in the units of the appropriate sensation continuum. By definition, all sensation jnd's for a given sensation continuum are equal to one another, given unchanged values of all stimulus properties except those on the primary stimulus continuum. A stimulus jnd is the magnitude of the change on the primary stimulus continuum, measured in appropriate physical units, which is just sufficient to produce a change of one sensation jnd upward at that point. (A discussion of the essentially statistical nature of jnd's appears later in this paper.) In general, stimulus jnd's will have different sizes at different points on the primary stimulus continuum. The rest of this paper will not be intelligible unless you keep the distinction between these two kinds of jnd's in mind.

We have assumed that jnd's are measured upward on the stimulus continuum. They could also be measured downward, and the possibility exists that the two measurements might not agree. In fact, they are certain not to agree if the distance spanned is more than two jnd's, and if the size of the jnd at the end where measurement starts is used as the unit of measurement, since this means that the size of the measurement unit will be different depending on direction of measurement. However, such discrepancies might exist in the measurement of a single jnd; this, if it happened, would mean that jnd's are not suitable units of measurement unless direction is specified. We have therefore confined ourselves to upward jnd's.

Now we can say exactly what this paper is about. Given a function (obtained from experiment, theory, or both) relating stimulus to sensation jnd's for all points of the primary stimulus continuum, what may we infer about the sensory scale implied by that jnd function?

Fechner's derivation of Fechner's law. On October 22, 1850, Fechner (2) thought up the first (incorrect) answer to the question which ended the previous paragraph. Let us call any function that gives the size of a stimulus jnd at each point of the stimulus continuum a *Weber function* (corresponding to "a function relating stimulus to sensation jnd's" of the previous section), and any one-to-one function based on cumulated jnd's which relates the stimulus continuum to a sensory scale a *Fechner function* (corresponding to "a sensation scale" of the previous section). These definitions do *not* restrict our attention to those two special functions which have come to be known in psychophysics as Weber's law and Fechner's law! Fechner believed that the Fechner function corresponding to any Weber function could be expressed as the solution (integral) of a first order linear differential equation involving that Weber function. He applied this procedure to Weber's Law, which asserts that for a given stimulus continuum the size of the stimulus jnd, Δx, divided by the value of the stimulus at that point, x, is a constant ($\Delta x/x = k$). Let us examine his argument.

If Weber's law is true, then, since all sensation jnd's are equal by definition, there is a constant A such that

$$\frac{\Delta u}{\Delta x} = \frac{A}{x} \qquad [1]$$

where Δu denotes the size of the sensation jnd. The heart of Fechner's solution to his and our basic problem was to "rewrite" Equation 1 as the differential equation

$$\frac{du}{dx} = \frac{A}{x} \qquad [2]$$

How did Fechner make this step from differences (deltas) to differentials? He used what he called a "mathematical auxiliary principle," the essence of which is that what is true for differences as small as jnd's ought also to be true for all smaller differences and so true in the limit as they approach zero (differentials). If this argument were acceptable (which it is not), the rest would be simple. Equation 2, when integrated, yields the familiar logarithmic relationship between sensation and stimulus which is known as Fechner's law.

Fechner thought that his general procedure ought to be applicable to any Weber function, not just to Weber's law. It is not. Except for a few special cases like Weber's law, the definition of sensation jnd's as equal and the "mathematical auxiliary principle" are mutually contradictory. For example, consider the Weber function $\Delta x/x^2 = k$. Then, following Fechner's procedure, we should write:

$$\frac{\Delta u}{\Delta x} = \frac{A}{x^2} \quad \text{and so} \quad \frac{du}{dx} = \frac{A}{x^2}$$

Integrating, we get

$$u = B - \frac{A}{x}$$

Let us now check to see whether this new Fechner function satisfies the definition which says that sensation jnd's are equal to one another. If we are at point x on the stimulus continuum, a stimulus jnd, according to the Weber function used in this example, is kx^2. The sensation increment corresponding to this change, the sensation jnd at this point, is therefore given by:

$$\begin{aligned} u(x + kx^2) - u(x) \\ &= B - \frac{A}{(x + kx^2)} - B + \frac{A}{x} \\ &= \frac{Ak}{1 + kx} \end{aligned}$$

which is clearly not a constant for any value of the constant A except zero.

This, although only one example, is typical in the sense that almost any example you could think of would show the same discrepancy. Only for a very few Weber functions—some pathological ones, Weber's law, and its generalization $\Delta x = kx + c$—does the "mathematical auxiliary principle" yield a Fechner function with equal jnd's. We will not take space to prove this formally, but a formal proof is available.

The functional equation solution. We have shown that Fechner's procedure involves a self-contradiction. We shall show later that it leads to wrong results in all important cases except Weber's law. Obviously the "mathematical auxiliary principle" is wrong and must go.

How, then, should we cumulate jnd's? The simplest, most obvious procedure (which has very often been used exactly because it is simplest and most obvious) is simply to add them up one at a time. If the first jnd on a primary stimulus continuum is 5 stimulus units, then two points on our cumulated jnd scale should be 0, 0 and 1, 5, where the first number is the scale value on the y axis and the second number is the corresponding stimulus value on the x axis. If we then find that the size of the stimulus jnd at 5 on the stimulus continuum is 8, then the third point is 2, 13. If we find that the size of the stimulus jnd at 13 is 10, then the fourth point is 3, 23, and so on.

Fechner and some of his more modern imitators went way out of their way to avoid this simple and sensible procedure; in retrospect it is hard to decide why they did so. At any rate, the next two sections of this paper will develop a formal mathematical solution to Fechner's mathematical problem—a solution which turns out to be the mathematical equivalent of the simple graphical or arithmetic technique discussed in the previous paragraph. The mathematical problem centers about how to fill in the curve between the discrete points arrived at by the graphical method.

What mathematical tools can we use to replace Fechner's "mathematical auxiliary principle"? Equation 1, and the corresponding ones based on other Weber functions, can be solved directly without any mathematical auxiliary principles or other further assumptions. They are examples of what mathematicians call functional equations. The papers on which most of our discussion is based (5, 6) were published in the 1880s, twenty years after Fechner first published his work.

The kind of functional equation implied by the definition of equality of sensation jnd's is soluble for a very wide class of Weber functions. Unfortunately, there is an infinity of inherently different solutions to each of these equations. However, further consideration of what we mean by a sensation scale will lead us to properties which we usually take for granted and which are enough to narrow the solutions down to just one interval scale, unique except for its zero point and unit of measurement. It is interesting that in the case of the linear generalization of Weber's law, and in that case only, the functional-equation solution is the same as that obtained by Fechner's auxiliary principle; for all other Weber functions the two solutions are different.

First, we will state the general mathematical problem and its solution. Let x, $x \geq 0$, denote a typical value of the stimulus continuum, and let u denote the (unknown) Fechner function. Let g be the (given) Weber function; i.e., a stimulus magnitude y, $y \geq x$, is detected as larger (in a statistical sense) than x if $y > x + g(x)$, whereas it is not discriminated as different from x if $x \leq y \leq x + g(x)$. We write $x + g(x) = f(x)$. By definition, a sensory jnd at the sensation $u(x)$ is given by the increment[2]

$$u[f(x)] - u(x)$$

(In the usual "delta" notation, $g(x) = \Delta x$ and $u[f(x)] - u(x) = \Delta u$.) The condition that sensation jnd's be equal simply means that all sensation jnd's are a constant, which we may take to be 1 for convenience, since an arbitrary change of unit does not matter. Thus, we have our major mathematical problem:

> Find those real-valued differentiable functions u, defined for all $x \geq 0$, such that $u[f(x)] - u(x) = 1$, for all $x \geq 0$.

Note that we have said *those* functions, not *that* function, for there may be more than one such function. This uniqueness question has not traditionally been raised, for so long as the problem was formulated in terms of linear differential equations, the uniqueness theorems of that branch of mathematics insured only one solution. In the realm of functional equations, we have no such assurances.

It is very lucky that the functional equation which has arisen in this problem is one of the more famous in the literature; it is called Abel's equation.[3] The principal results we shall need concerning this equation were presented by Koenigs (**5**, **6**) in 1884 and 1885.[4] First, we will present his uniqueness results, which illustrate the method of attack and lead up to the general solution. Suppose that $u_0(x)$ is a solution to Abel's equation, and suppose $p(x)$ is an arbitrary periodic function with period 1—in other words, any function satisfying

---

[2] Throughout this paper we shall have to use functions of functions. In general, if v and w are two real-valued functions of a real variable x, $v[w(x)]$ denotes the number obtained by calculating $y = w(x)$ and then finding $v(y)$. Clearly, the order of writing v and w is material, for $v[w(x)]$ does not generally equal $w[v(x)]$. Consider, for example, $v(x) = ax$, where $a \neq 1$, and $w(x) = x^2$. Then, $v[w(x)] = v(x^2) = ax^2$, whereas $w[v(x)] = w(ax) = a^2x^2$.

$$p(x+1) = p(x)$$

$K \sin 2\pi x$ is periodic with period 1, and so is an example of a function $p(x)$. It is easly to show that the function $u_p(x) = u_0(x) + p[u_0(x)]$ is also a solution to Abel's equation:

$$\begin{aligned} u_p[f(x)] &= u_0[f(x)] + p\{u_0[f(x)]\} \\ &= 1 + u_0(x) + p[1 + u_0(x)] \\ &= 1 + u_0(x) + p[u_0(x)] \\ &= 1 + u_p(x) \end{aligned}$$

Furthermore, it can be shown that if u and u^* are two solutions to Abel's equation, then there exists a periodic function p with period 1 such that

$$u(x) = u^*(x) + p[u^*(x)]$$

Thus, if we have any solution u_0 to our problem and if we choose p to be a differentiable periodic function with period 1, then $u_p = u_0 + p(u_0)$ is also differentiable and solves the problem.

In the case of Weber's law, we have $f(x) = kx$, $k > 1$, and the differentiable function $u_0(x) = \dfrac{\log x}{\log k}$ is easily shown to satisfy the condition of equal sensation jnd's. Therefore,

$$\frac{\log x}{\log k} + p\left(\frac{\log x}{\log k}\right)$$

is also a solution if p is differentiable and periodic with period 1.

There is an infinity of such functions p, and so in infinity of different solutions to the problem for any Weber function, including Weber's law. This, of course, is quite unsatisfactory; later on we will show that one of the properties that we usually attribute to jnd's, and which as yet we have not used, enables us to insure a unique solution. However, first it will be useful to present Koenigs's results on the existence of solutions to Abel's equation.

The existence of solutions to Abel's equation. In psychophysical problems, there is always a threshold $R > 0$, such that $g(x)$ is not observable in the range $0 \le x \le R$. Thus, it is only a matter of convenience what we assume about the behavior of g near 0; we shall suppose that

$$g(0) = 0 \quad \text{and} \quad 0 < g'(0) < \infty$$

where $g'(x) = \dfrac{dg}{dx}$. It is known also from experimental work that g is never 0 and that on the whole it will increase with x, except for limited ranges of some stimuli, where it may decrease slowly. With little or no loss of generality, we may suppose it never decreases so rapidly as to have a slope less than -1. In other words, we also assume:

$$g(x) > 0 \quad \text{and} \quad g'(x) > -1 \quad \text{for} \quad x > 0$$

From these assumptions, it follows that $f(x) = x + g(x)$has these properties:

---

[3] Sometimes this equation is spoken of as the Abel-Schroder equation, but more often Abel's name is attached to this equation and Schroder's name to the equation $v[f(x)] = cv(x)$, which arises from Abel's equation through the substitution $v = c^u$.

[4] We are indebted to Richard Bellman of the RAND Corportion for directing us to the literature on the Abel equation.

> f if strictly monotonic in x, i.e., if $x < y$, then $f(x) < f(y)$; 0 is the only fixed point of f (x is a fixed point if $f(x) = x$); and $1 < f'(0) < \infty$.

The strict monotonicity of f implies that there exists an inverse function f^{-1}, i.e., a function such that

$$f^{-1}[f(x)] = x = f[f^{-1}(x)]$$

It is easy to show that:

> f^{-1} is strictly monotonic increasing, x is a fixed point of f^{-1} if and only if $x = 0$, $0 < f^{-1\prime}(0) < 1$.

Observe that if we know a solution v to the equation

$$v[f^{-1}(x)] = 1 + v(x) \qquad [3]$$

then $u = -v$ is a solution to

$$u[f(x)] = 1 + u(x) \qquad [4]$$

So it will suffice to deal with f^{-1}. If, in addition to the three properties mentioned, f^{-1} is analytic, i.e., if there exist constants a_i such that

$$f^{-1}(x) = \sum_{i=0}^{\infty} a_i x^i,$$

then Koenigs has shown that a differentiable solution exists to Abel's equation. In applications, analyticity is no real restriction. For simplicity of notation, let us denote f^{-1} by h; then Koenigs' theorem (which is not easy to prove) may be expressed as follows: Let $h^{(n)}$ denote the n^{th} iterate of h (i.e., $h^{(n)}(x)$ is the result of n successive applications of h beginning at the point x), and let

$$\phi(x) = \lim_{n\to\infty} \frac{h^{(n)}(x)}{[h'(0)]^n}$$

then ϕ exists and is differentiable, and

$$v_0(x) = \frac{\log \phi(x)}{\log h'(0)}$$

is a solution to Abel's Equation 3. Therefore, since $h'(0) = 1/f'(0)$,

$$u_0(x) = \frac{\log \phi(x)}{\log f'(0)}$$

is a solution to Equation 4 and so to our problem.

The difficult part of the proof is to show that the limit exists. Assuming that it does, it is easy to show that $u_0(x)$ is a solution. Since $h[f(x)] = x$, $u_0[f(x)]$

$$= \left\{\log \lim_{n\to\infty} \frac{h^{(n)}[f(x)]}{[h'(0)]^n}\right\} \Big/ \log f'(0)$$

$$= \left\{\log \frac{1}{h'(0)} \lim_{n\to\infty} \times \frac{h^{(n-1)}(x)}{[h'(0)]^{n-1}}\right\} \Big/ \log f'(0)$$

$$= \frac{\log f'(0) + \log \lim_{n\to\infty} \frac{h^{(n-1)}(x)}{[h'(0)]^{n-1}}}{\log f'(0)}$$

$$= 1 + u_0(x)$$

The evaluation of the above limit for ϕ is rarely a simple task. Furthermore, the conditions under which it has been shown to exist and to provide a solution to Abel's equation are only sufficient conditions—there are other circumstances in which solutions exist. For example, the function $f(x) = ax^b$, $b \neq 1$, fails to satisfy $1 < f'(0) < \infty$, yet by direct verification one can show that

$$u_0(x) = \frac{\log \log [a^{1/(b-1)}x]}{\log b}$$

satisfies $u_0(ax^b) = 1 + u_0(x)$. The function $f(x) = x + ax^b$ also fails to meet the same condition, but a solution probably exists in this case too. Presumably, other functions can be found which approximate empirical data and which meet the assumed conditions, but it remains to be seen whether the limit ϕ can be evaluated.

The difficulty is, first, in inverting f, and second, in finding a simple expression for $h^{(n)}$. Since this is generally difficult, we doubt that the mathematics of this section will be useful to psychophysicists who want a nongraphic method for cumulating jnd's.

It should be pointed out again that for the empirically important Weber function $g(x) = kx + c$ the solution is known: it is

$$u_0(x) = \frac{\log (kx + c)}{\log (1 + k)}$$

A further definition of the sensation continuum. So far we have examined two formulations of Fechner's problem, both of which are unsatisfactory. The first, that of Fechner, contains an internal contradiction. The second, the functional equation formulation, we have shown can be solved. Unfortunately, we have also shown that it has infinitely many families of different solutions, which is intolerable. In this section we shall propose an addition to the second formulation which amounts to a method of summating jnd's. We shall show that if we demand a particular form of invariance of distances measured in jnd units, then there is a unique (except for zero and unit) sensation scale for each of a wide variety of Weber functions, and for Weber's law this sensation scale is Fechner's law.

The common psychological custom for measuring distances in jnd's between two points is to use the size of the jnd at the lower point as the unit of measurement. Although it is rarely if ever explicitly stated, it is certainly implicitly assumed that if the distances ab and cd are both α stimulus jnd's in length, then they have an equal number, say $K(\alpha)$, of sensation jnd's. As a formal mathematical condition, this states that

$$u[x + \alpha g(x)] - u(x) = K(\alpha)$$

where K is some fixed, but unknown, function of α. It can be shown, first, that if u is a solution to this problem, then it must be the integral $\int \frac{dx}{g(x)}$ given by Fechner, but, second, that there are no solutions except when $g(x) = cx$ (Weber's law). We will not present a proof of this result since it is a blind alley, but we believe that it suggests that this customary measurement of distances should be abandoned.

We must now consider how such distances really should be measured. If x and y are more than one jnd apart, we may expect the size of the jnd to change as we go from x to y. That fact should be taken into account in using jnd's as units of measurement; failure to take it into account is what makes Fechner's auxiliary principle and the standard measuring procedure unacceptable. We shall proceed to formulate this more sensible method of using jnd's as measuring units.

Let $f(x) = x + g(x)$; then the point $f(x)$ is one x-jnd larger than x. The point $f[f(x)] = f^{(2)}(x)$ is one $f(x)$-jnd larger than $f(x)$. In general $f^{(n)}(x)$ is one $f^{(n-1)}(x)$-jnd larger than the point $f^{(n-1)}(x)$. Clearly, for $y > x$, we can find some integer n such that

$$f^{(n)}(x) \leq y < f^{(n+1)}(x)$$

and it is reasonable to say that y is between n and $n + 1$ jnd's larger than x. For the moment, let us suppose that y was chosen so that $y = f^{(n)}(x)$, then we can say y is exactly n jnd's larger than x. It seems plausible to require that the same be true of the sensory continuum, i.e.,

$$u[f^{(n)}(x)] - u(x) = n$$

In words, we are saying that if point y is 20 stimulus jnd's higher than point x on the stimulus continuum, then it must also be 20 sensation jnd's higher than point x on the sensation continuum. If the above condition is met for $n = 1$ (in other words, if all sensation jnd's for a given sensory continuum are equal), then it must also be met for all larger values of n, since

$$\begin{aligned} u[f^{(n)}(x)] &- u(x) \\ &= u\{f[f^{(n-1)}(x)]\} - u(x) \\ &= 1 + u[f^{(n-1)}(x)] - u(x) \\ &\dots \\ &= n \end{aligned}$$

But this takes care of relatively few points, and does not allow us to say exactly how many jnd's y is from x unless the difference is a whole number of jnd's. We must find a definition which tells us how to subdivide a jnd into fractional parts. How to do this is not obvious, since the definition of distances given above involves iterates of f, and these are apparently defined only for integers. Fortunately, it is possible to generalize the notion of an iterate to arbitrary, rather than integral, indices. This problem is closely related to that of Abel's functional equation which Koenigs examined; we shall be able to use his results.

First, we can set up some properties that a generalized iterate $f^{(t)}(x)$, where t is any non-negative number, should meet. In essence, they amount to stipulating that $f^{(t)}(x)$ should coincide with the usual definition when t is an integer and that the same law of composition should hold. Formally, it is sufficient to require that

$$f^{(0)}(x) = x, \quad f^{(1)}(x) = f(x)$$

and for every s and $t \geq 0$,

$$f^{(s+t)}(x) = f^{(s)}[f^{(t)}(x)]$$

For integers, the generalized iterate coincides with the usual notion, as you can see, by repeatedly applying the last condition to the second one.

We have already presented a result of Koenigs which showed that if f is strictly monotonic and analytic, $1 < f'(0) < \infty$, and 0 is the only fixed point of f, then there exists a function ϕ defined in terms of the iterates of f^{-1} such that

$$u_0(x) = \frac{\log \phi(x)}{\log f'(0)}$$

is a basic solution to Abel's equation. This means that ϕ is itself a solution to what is called Schroder's equation

$$v[f(x)] = f'(0)v(x)$$

which is obtained from Abel's by taking exponentials on both sides. Using this fact and following Koenigs, it is easy to show that ϕ^{-1} exists and that the function

$$f^{(t)}(x) = \phi^{-1}\{[f'(0)]^t\phi(x)\}$$

satisfies the three conditions of a generalized iterate. We show the latter. First,

$$f^{(0)}(x) = \phi^{-1}[\phi(x)] = x$$

Second,

$$f^{(1)}(x) = \phi^{-1}[f'(0)\phi(x)]$$

And so, using the fact that ϕ satisfies Schroder's equation,

$$\phi[f^{(1)}(x)] = f'(0)\phi(x) = \phi[f(x)]$$

Hence,

$$f^{(1)}(x) = f(x)$$

Finally,

$$\begin{aligned} f^{(s)}&[f^{(t)}(x)] \\ &= \phi^{-1}[[f'(0)]^s\phi(\phi^{-1}\{[f'(0)]^t\phi(x)\})] \\ &= \phi^{-1}\{[f'(0)]^s[f'(0)]^t\phi(x)\} \\ &= \phi^{-1}\{[f'(0)]^{s+t}\phi(x)\} \\ &= f^{(s+t)}(x) \end{aligned}$$

So, with this definition of the generalized iterate, we can generalize the above definition of distances in jnd's to prescribe how to deal with fractional jnd's.

We reformulate our major mathematical problem:

> Given a Weber function g which is analytic, $g'(x) > -1$ for all $x > 0$, $g'(0) > 0$, and $g(0) = 0$, to find those functions $u(x)$ such that $u[f^{(t)}(x)] - u(x) = t$, for all $x > 0$, and all $t > 0$, where $f^{(t)}$ is the generalized iterate of $f(x) = x + g(x)$.

Note that by setting $t = 1$, this condition implies the equality of sensation jnd's.

First, we show that $u_0(x) = \dfrac{\log \phi(x)}{\log f'(0)}$

solves the reformulated problem:

$$\begin{aligned} u_0[f^{(t)}(x)] - u_0(x) &= \frac{\log \phi[f^{(t)}(x)]}{\log f'(0)} - \frac{\log \phi(x)}{\log f'(0)} \\ &= \frac{\log \{[f'(0)]^t \phi(x)\} - \log \phi(x)}{\log f'(0)} \\ &= \frac{t \log f'(0)}{\log f'(0)} \\ &= t \end{aligned}$$

Second, from the results about Abel's equation, we know that if there are any other solutions to this problem, they must be of the form $u_p = u_0 + p(u_0)$, where p is periodic with period 1. For u_p actually to solve the reformulated problem, it is necessary that for every $t \geq 0$,

$$\begin{aligned} t &= u_p[f^{(t)}(x)] - u_p(x) \\ &= u_0[f^{(t)}(x)] + p\{u_0[f^{(t)}(x)]\} \\ &\qquad - u_0(x) - p[u_0(x)] \\ &= t + p[t + u_0(x)] - p[u_0(x)] \end{aligned}$$

Thus, for every $t \geq 0$,

$$p[t + u_0(x)] = p[u_0(x)]$$

That is, p must be periodic with every period t, and so p is a constant. Thus, up to an additive constant, u_0 is the unique function which solves our reformulated problem.

In nonmathematical language, introducing the method of measuring fractional jnd's has enabled us to eliminate all solutions to Abel's equation save u_0, thus cutting down the number of acceptable solutions from infinity to one.

We conclude, therefore, that the condition stated in the reformulated problem constitutes an acceptable definition of a psychophysical sensation continuum, in the sense that it yields a unique Fechner function for any reasonable Weber function. We also find that for Weber's law this condition yields Fechner's law. The solution of our reformulated problem may cause unhappiness because it is not the same as the integral "solution" proposed by Fechner, except in the special case of the linear generalization of Weber's law. However, we have already shown that the integral "solution" contradicts the definition of equal sensation jnd's.

It is sad that the integral is not the right solution, for its evaluation is often easy, and we fear that no working psychophysicist will find in our mathematics a tool for determing a summated jnd scale any better or more efficient than the simple graphic procedure of adding jnd's up one at a time.

The statistical nature of jnd's. So far we have sounded as though we were treating jnd's as fixed quantities, although every psychophysicist knows that jnd's are statistical fictions, defined by an arbitrarily chosen cutoff on a cumulative frequency curve.

However, we now show that our method of reducing the infinity of solutions to Abel's equation to one is equivalent to treating jnd's as just such statistical fictions.

We start with the old, famous psychological rule of thumb: equally often noticed differences are equal, unless always or never noticed. We define $P(y,x)$ as the probability that y is discriminated as larger than x. Now, this rule of thumb simply means that on the sensation continuum the function $P(y,x)$ is transformed in such a way that it no longer depends on x and y separately, but only on the difference of their transformed values. Put another way, the subjective continuum u is a strictly monotonic transformation of the stimulus continuum such that the probability that a change of δ units on the sensation scale will be detected depends only upon δ, and not on the place at which δ begins or ends.

Formally, if we are at a point x of the stimulus continuum, and therefore at $u(x)$ on the sensation scale, and if a stimulus y is presented such that $u(y) = u(x) + \delta$, then the chance that y will be detected depends upon δ, but not on x. If we note that

$$y = u^{-1}[u(x) + \delta]$$

then the condition is that

$$P\{u^{-1}[u(x) + \delta], x\} = P(\delta)$$

Our problem is to decide under what conditions this problem has a solution and what that solution is. To this end, we make the assumption that for each x, $P(y,x)$ is a strictly monotonic increasing function of y.

We show the following: If the above problem has a solution, then there exists a function $f(x)$ such that $P[f^{(t)}(x), x]$ is independent of x, where $f^{(t)}(x)$ is the t^{th} iterate of $f(x)$ previously defined. The function $f(x) - x$ is a Weber function naturally defined in terms of P. If there is a solution, it is unique and it is the solution u_0 to Abel's equation $u[f(x)] - u(x) = 1$. In other words, if there is any solution to the problem of equally often noticed differences being equal, then it is unique and it is the solution to our proposed reformulation of Fechner's problem.

The proof is comparatively simple and runs as follows. Suppose there exists a solution u to the condition that $P\{u^{-1}[u(x) + \delta], x\}$ is independent of x for all $\delta \geq 0$. Since P is strictly montonic in y for all x, there is a unique solution to $P(y,x) = k$ for each k, $0 < k < 1$; call it $y = f_k(x)$. For any δ, let $k = P(\delta)$, and so by our assumption u must satisfy

$$u^{-1}[u(x) + \delta] = f_\delta(x)$$

where we have written f_δ for $f_{P(\delta)}$. Applying u to this, we have

$$u[f_\delta(x)] - u(x) = \delta$$

Let $f = f_1$. We observe that if $\delta = 0$, then $f_0(x) = x$. Suppose we choose any $\delta, \epsilon \geq 0$ and let $y = f_\epsilon(x)$; then

$$\begin{aligned} \delta &= u[f_\delta(y)] - u(y) \\ &= u\{f_\delta[f_\epsilon(x)]\} - u[f_\epsilon(x)] \\ &= u\{f_\delta[f_\epsilon(x)]\} - u(x) - \epsilon \end{aligned}$$

Thus,

$$u\{f_\delta[f_\epsilon(x)]\} - u(x) = \delta + \epsilon$$

But, from above,

$$u[f_{\delta+\epsilon}(x)] - u(x) = \delta + \epsilon$$

so

$$u[f_{\delta+\epsilon}(x)] = u\{f_\delta[f_\epsilon(x)]\}$$

whence

$$f_{\delta+\epsilon}(x) = f_\delta[f_\epsilon(x)]$$

Thus, we have shown that f_δ must satisfy the three conditions of a generalized iterate of f, i.e., $f_\delta = f^{(\delta)}$ for all δ, so a necessary condition for a solution is that

$$P[f^{(\delta)}(x), x]$$

shall be independent of x. From the fact that $u[f_\delta(x)] - u(x) = \delta = u[f^{(\delta)}(x)] - u(x)$, it follows that the solution is uniqe and that it is the same as that given for our reformulation of Fechner's problem.

It probably is not obvious, but the point of this section extends beyond sensory psychophysics into the scaling procedures based on Thurstone's law of comparative judgment. Case V of that law is based on the assumption that equally often noticed differences are equal unless always or never noticed. This fact has two interesting implications. The first and more obvious one is that these two apparently different branches of psychoological measurements are actually doing the same thing (namely, using a measure of confusion as a unit of measurement by assuming that confusion is equal at all places on the subjective scale). The second, less obvious implication is that perhaps sensory psychophysics can profit by considering, as Thurstone and his followers have, scaling methods with less rigid assumptions which nevertheless are based on confusability data. One of us (Luce) will pursue this possibility further in a forthcoming book (7).

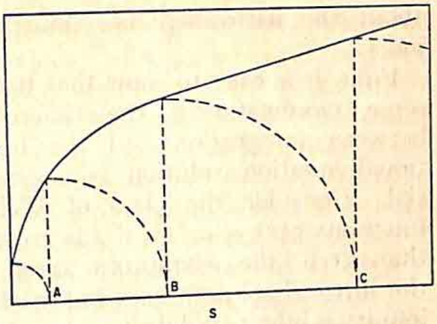


FIG. 1. How to cumulate jnd's. The size of the jnd at the origin is marked off on the x axis to find point A, the size of the jnd at A is marked off to find point B, and so on. The stimulus values A, B, C, . . . correspond to the points 1, 2, 3, . . . on the cumulated jnd scale.

Graphic methods for cumulating jnd's. Psychophysical data do not come in mathematical form. In order to apply our method for cumulating jnd's (or Fechner's, for that matter), it is necessary either to put the Weber function into equation form, or else to develop a graphic equivalent of the appropriate mathematical operations. The graphical equivalent of Fechner's technique is well known, although rarely used (see, e.g., **15**, pp. 94 and 147–148). It is, of course, wrong, since Fechner's technique is wrong. If our technique is to be of greatest applicability, we should provide a graphic equivalent also. Unfortunately, it seems difficult to find a truly convenient one. The only method we know of is to go back to the basic idea of adding up jnd's—the idea that one jnd plus one jnd is two jnd's. The method of applying this basic idea is given in Figure 1, and was discussed earlier in the paper. Its error characteristics are abour the same as those of the graphic techniques of integration which have been used in the past. Unfortunately, the method is tedious; if there are 170 jnd's between absolute threshold and the upper limit of discrimination, then 170 separate operations are required to determine the cumulated-jnd scale. The errors in these successive operations do not multiply, however.

Practial effects of the new procedure. No doubt it is important to understand Fechner's logical error and to know how to avoid it, but the burning question for working psychophysicists is: What, if anything, does this do to the currently accepted conclusions about the uselessness of adding up jnd's?

First, it is easy to show that under some circumstances the difference between integration and the functional-equation solution is substantial. Consider the class of Weber functions $g(x) = ax^{1+e}$: if e is greater than zero, the asymptotic error of the integral solution as x approaches infinity is infinite; while if e is less than zero, the asymptotic error is zero. Of course, if e equals zero (Weber's law), the two procedures give identical results. The order of magnitude of the error for small numbers of jnd's depends on the constants in the equation; it can be of significant size even if e is less than zero. One way of looking at it is that the integral solution is the approximation given by the first two terms of a Taylor series expansion of the functional equation; all square and higher power terms of the expansion are omitted:

$$\begin{aligned} u[x + g(x)] - u(x) &= 1 \\ &= u(x) + u'(x)g(x) \\ &\qquad + u''(x)\frac{g(x)^2}{2!} + \dots - u(x) \\ &= u'(x)g(x) \\ &\qquad + \left\{u''(x)\frac{g(x)^2}{2!} + \dots\right\} \end{aligned}$$

A number of experimental determinations of jnd's, particularly for intensive continua, produce a curve of $g(x)/x$ that first falls and then is flat—a function often well approximated by $g(x) = kx + c$. However, for some continua the picture is less simple. There are some (pitch, for example) where the curve appears to rise again at the high end. The falling section of these curves corresponds to the case $e < 0$; the flat section corresponds to the case $e = 0$; the rising section corresponds to the case $e > 0$. However, the x-axis of such graphs is usually plotted logarithmically. This means that the rising section may cover most of the range within which the stimulus can be varied—a fact which the logarithmic x-axis tends to conceal. So it is quite possible that the error in using the integration technique is substantial for many sense modalities and for large ranges within each.

But the possibility of error is irrelevant unless someone has actually made the error. Has anyone? Extensive examination of the literature suggests that the answer is that not very many such errors have occurred. Some authors are quite unclear about how they added up jnd's, but many of them have preferred the step-by-step method which corresponds to the functional-equation solution because it was very simple to do. How simple it is, of course, depends on the number of jnd's to be added; we doubt very much if the jnd's for pitch will ever be added this way, since there are several thousand of them. We have found only one clear instance (15) in which the graphic equivalent of integration has been used (to cumulate pitch jnd's, as it happens), though it has been vigorously recommended. The general avoidance of the graphic equivalent of integration may be caused by shrewd intuition that something is wrong with Fechner's mathematical auxiliary principle. Or it may simply be a rare instance in which the fear of mathematical complexity has benefited science.

Do cumulated jnd's agree with other scales? The results of cumulating jnd's have often been compared with the results of other psychophysical procedures (4). The most common finding has been that the cumulated jnd scales do not agree with scales determined by fractionation or direct magnitude estimation, at least for such continua as loudness. A review of this literature might seem appropriate here, but it is quite *unnecessary*, since the relation between scales based on confusion data (like cumulated jnd scales) and those based on fractionation or magnitude estimation has been extensively and excellently discussed in recent studies by Stevens (14), Stevens and Galanter (16), and Piéron (8, 9, 10).

The controversy over the relation between cumulated jnd scales and scales determined by other methods is embedded in a larger, sometimes acrimonious controversy about the relationships among various methods of sensory scaling. To some extent we shall have to enter the fray.

The first and most important question is this: Do the different scaling procedures, if properly used, lead to different scales? Unless we reject a great many experiments as improperly performed, we must answer "Yes." But the issue is not as simple or unambiguous as that answer. For example, Garner (3) has developed a loudness scale based on both fractionation and multisection judgments that fits a large number of experimental results in auditory psychophysics better than does the old sone scale (his paper was written prior to the development of the new sone scale [13]). Figure 2 shows the relationship between that scale and a cumulated jnd scale for loudness prepared by us from Riesz's data (11, 12). The two scales seem to be roughly linearly related—but does it mean anything for the controversy? Riesz's procedure has often been criticized, and his data are almost 30 years old. The form of Garner's scale (which is all that matters for this argument) is based primarily on his multisection rather than his fractionation data. Scales based upon multisection data usually agree with those constructed by confusability methods; the explanation proposed by critics of these methods is that the adjustment of five or six stimuli in a multisection experiment may produce confusion among the tones being adjusted. If this argument is correct, and if the form of Garner's scale is based upon multisection data, it is not surprising that the two agree. Our reason for so extensive a discussion of Garner's scale is that loudness is the central battleground of this controversy: If the verdict of psychophysical history is that confusability and multisection scales give results different from fractionation results for loudness, then psychologists will almost certainly assume that the two procedures yield different results in other intensive (or, as Stevens calls them, Class I or prothetic) continua. Unfortunately, even in psychophysics, not enough universally accepted data are available to settle the argument.

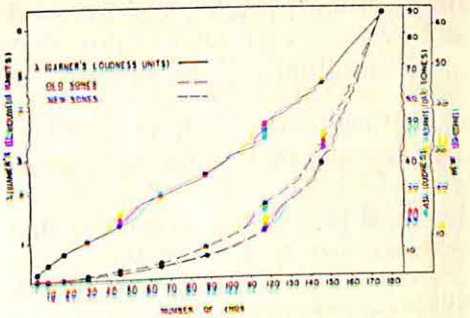


FIG. 2. The relation between Garner's loudness scale and Riesz's cumulated jnd scale. The old sone scale (ASA loudness scale) and Stevens's recent revision of it are included for comparison.

If confusability scales and scales based upon fractionation or direct magnitude estimation agree, no problem arises. If not (and we suspect they will not), psychophysicists must still evaluate each kind of procedure and its resulting scale. Some psychophysicists feel that fractionation and magnitude estimation have great face validity, and that confusability scales are distortions of the scales obtained by these procedures. They say that fractionation and estimation scales correspond to what *S*s say they feel, they are obtained by straightforward procedures rather than indirect ones, and, after all, what logic is there in basing a measure of magnitude on variance or "noise."

Other psychophysicists feel that confusability scaling is the better method. They say that fractionation and estimation data are unreliable, variable, and, as a rule, at least fractionation data cannot be turned into scales unless obtained from a "good," which means extensively trained, subject. The estimation techniques have not been used enough times in enough places to indicate clearly what effect, if any, training may have on the results. Confusability scales can be obtained from untrained *S*s who have no idea what form of scale is wanted from them; they can even be obtained from animals.

Each group asserts that its preferred scales are more nearly consistent with the bulk of psychophysical data than the other kind of scales; each group can produce impressive arguments to buttress its claim.

Still another position is possible: perhaps two different kinds of sensory processes are being tapped by these two different kinds of procedures. If so, both kinds of scales are useful, but for different purposes. This could well be the eventual endpoint of the argument.

Yet another source of confusion in the argument is the treatment of individual differences. The custom has been to take means or medians, and recently a number of psychophysicists have raised vigorous questions about the appropriateness of doing so. W. J. McGill[5] is currently attempting to find a better way of respecting individual differences while still obtaining a "universal" scale. It will be interesting to see what light serious attempts to do justice to individual differences sheds on the differences between the two classes of scales.

The status of cumulated jnd's has been controversial for more than a hundred years, and this paper is not intended as an attempt to settle the controversy. Our main point is that Fechner's problem has been improperly formulated and that the integral usually offered as a solution is not in fact a solution when the Weber function differs from $g(x) = kx + c$. We have also developed what appears to be the correct solution, only to find that in computational work it has usually been used in spite of its disagreement with the integral solution. This means that our clarification of the logical issues underlying Fechner's formulation does little to change the status of the present, primarily empirical, controversy about scaling methods. However, one of us (Luce [7]) has recently developed a way of dealing with confusability data based on a simple axiom which, if it works out successfully, may resolve the difficulty by changing our ideas about the meaning of confusability scales; this development will be described in another publication.

SUMMARY

Fechner's method for adding up just noticeable differences (jnd's) to obtain sensory scales is based on a

---

[5] W. J. McGill, Personal communication, 1957.

mathematical error: he used a differential equation approximation to a functional equation instead of the functional equation itself. The functional equation can, however, be solved directly. The solution coincides with the differential equation solution only in the special case in which the linear generalization of Weber's law holds exactly. The mathematical properties of the formal solution are such that it probably will not be very useful for practical computation, but the extremely simple graphical procedure of adding up jnd's one at a time is the graphical equivalent of the mathematically correct solution. The amount of difference between the two procedures can be calculated for some special cases; its size depends on the form of the function relating size of jnd's to stimulus magnitude.

This error does not seem to have any significant impact upon the controversy over the relation between cumulated jnd scales and scales based on fractionation and direct estimation data because most psychophysicists have, in fact, ignored the recommended (incorrect) procedure and have stubbornly summated jnd's in the obvious and correct way.

REFERENCES

1. BORING, E. G. *A history of experimental psychology.* (2nd ed.) New York: Appleton-Century-Crofts, 1950.
2. FECHNER, G. T. *Elemente der Psychophysik.* (Reprint) Leipzig: Breitkopf und Härtel, 1889.
3. GARNER, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1952, **24**, 153–157.
4. HALSEY, R. M., & CHAPANIS, A. Luminance of equally bright colors. *J. optical Soc. Amer.*, 1955, **45**, 1–6.
5. KOENIGS, M. G. Recherches sur les intégrales de certaines équations fonctionnelles. *Ann. Scientifiques de l'Ecole Normale Supérieure* (3), 1884, **1**, Supplement, S1-S41.
6. KOENIGS, M. G. Nouvelles recherches sur les équations fonctionnelles. *Ann. Scientifiques de l'Ecole Normale Supérieure* (3), 1885, **2**, 385–404.
7. LUCE, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, in press.
8. PIÉRON, H. L'évaluation des sensations. *Bull. Psychol., Univer. Paris*, 1950, **4**, 3–38.
9. PIÉRON, H. *Les problèmes fondamentaux de la psychophysique dans la science actuelle.* Paris: Hermann, 1951.
10. PIÉRON, H. *The sensations, their functions, processes and mechanisms.* New Haven: Yale Univer. Press, 1952.
11. RIESZ, R. R. The differential intensity sensitivity of the ear for pure tones. *Physical Rev.*, 1928, **31**, 867–875.
12. RIESZ, R. R. The relationship between loudness and the minimum perceptible increment of intensity. *J. accoust. Soc. Amer.*, 1933, **4**, 211–216.
13. STEVENS, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
14. STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.
15. STEVENS, S. S., & DAVIS, H. *Hearing: Its psychology and physiology.* New York: Wiley, 1938.
16. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received November 27, 1957)

Psychological Review
Vol. 65, No. 4, 1958

STRENGTH OF CARDIAC CONDITIONED RESPONSES WITH VARYING UNCONDITIONED STIMULUS DURATIONS[1]

NORMA WEGNER[2] AND DAVID ZEAMAN

University of Connecticut

The effects of long and short shocks on the conditioning of fear have been frequently considered important tests of theory of how fear is learned (1, 2, 3, 4, 9, 10). N. E. Miller, for example, has stated that according to a strict drive-reduction theory of reinforcement ". . . other things equal, a signal followed by a brief noxious stimulus should acquire the capacity to elicit stronger fear than one followed by a prolonged noxious stimulus" (2, p. 375).

Despite its theoretical significance, there have been few direct experimental attacks on this problem. We have found only two published conditioning studies with duration of noxious US as a major variable, one by Bitterman, Reed and Krauskopf (1) and the other by Mowrer and Solomon (4). Each of these studies examined only two values of shock duration over a relatively narrow range.

More extensive data on this problem can be extracted from a series of cardiac conditioning experiments we have done exploring the effects of duration of shock on the form of the conditioned heart response (8, 9, 10). These studies provide us with an assemblage of data, as yet unreported, on the relation of shock duration to strength rather than form of CR.

If heart rate change during anticipation of a shock is accepted as an index of fear, our classical conditioning studies using shocks of 0.1, 2.0, 6.0, and 15 sec. allow us to make a direct and thorough empirical check on Miller's statement of the drive reduction view. We shall first summarize our own experiments.

METHOD

General

The apparatus, procedure, and *S*s have all been described in detail in previous papers (8, 9, 10). They are summarized as follows. Forty-three male and twenty female college student *S*s were assigned unsystematically to four groups, each run under a classical trace-conditioning procedure with a different shock US duration (0.1, 2.0, 6.0, and 15 sec.). All groups received initially at least ten well distributed preconditioning trials of the CS alone—a 1-sec. tone (60 db, 512 c.p.s.). Following these were at least ten spaced conditioning trials consisting of the tone CS followed in 6 sec. by the shock US (13 V.A.C.). This mild but reportedly unpleasant shock was applied across the first two fingers of the left hand. Heart rate was measured with an electrocardiograph before, during, and after each trial.

Measures of CR and UR

Since individual differences in the form of conditioned cardiac responses occur, a measure of heart rate disturbance was chosen which would be comparable for the different types of response. This was a variability measure, the statistical *range* of heart rates during the conditioning (posttone) period. It was simply derived as follows. For a single *S* on every conditioning trial each heart beat interval between the tone and shock was converted to a rate measure. There were usually about eight or nine of these. The difference between the largest and smallest of them defined the *range* for that trial, a measure which we previously have designated as *Maximum CR*. A mean is taken over the first ten conditioning trials for this measure to represent a single *S*.

The same steps as above are taken to compute the *Maximum UR* except that ten beats

---

[1] This is Technical Report No. 15 under Contract Nonr-631 (00) between the Office of Naval Research and the University of Connecticut. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the University of Maryland.

after the shock onset are used instead of the eight or nine before.

RESULTS

Figure 1 presents the Maximum CR and Maximum UR for four groups having different shock durations. Each point is the mean of 15 *S*s' data (3 *S*s, selected without bias, were omitted to equalize the *N*s in the groups for statistical convenience). Maximum CR does not seem to vary systematically as a function of shock US duration. What fluctuations do occur are easily attributable to the corresponding variation in Maximum UR. This fact is established by Table 1, which gives the results of an analysis of covariance of the Maximum CR function, using Maximum UR as the relevant covariate. When the influence of the unconditioned response is extracted from the conditioned response there is left an insignificant amount of CR variation among the shock conditions.

The absence of significant differences among the group means is not to be interpreted as due to absence of conditioning. Evidence has already been presented (8, 9, 10) to show that a significant amount of cardiac conditioning took place in all four groups, using other (correlated) measures of response. To demonstrate further that the CRs and URs in Fig. 1 represent reliable disturbances in rate, we have compared them with similar measures of pretone variability, i.e., the range of ten pretone rate measures. Table 2 presents the results of a series of *t* tests of the significance of the differences between mean pretone range and mean posttone range (for CR), and between mean pretone range and postshock (for UR). Each *t* is based on 15 differences, each difference itself a mean (over 10 trials) for each of 15 *S*s. Pretone means were small, less than 1.5 B/M for all groups, in comparison to the posttone and postshock means shown in Fig. 1.

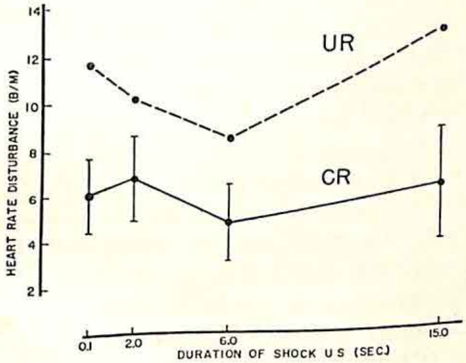


FIG. 1. The maximum amplitudes (range) of CR and UR are plotted as a function of shock US duration. One standard deviation is indicated by a vertical line on either side of each of the mean values comprising the points of the CR function.

TABLE 1

SUMMARY OF ANALYSIS OF COVARIANCE OF MAXIMUM CR AMPLITUDE WITH MAXIMUM UR AMPLITUDE

Source	*df*	M^2	*F*
Between durations	3	766	.68
Within	52	1131	

For 3 and 52 *df*, *F* is 2.79 for $p = 5\%$.
The variate-covariate correlation is +.57.

DISCUSSION

We wish to discuss our results briefly in relation (*a*) to previous findings and (*b*) to Miller's theoretical statement.

TABLE 2

PROBABILITY OF CHANCE OCCURRENCE OF MEAN DIFFERENCES BETWEEN PRETONE CARDIAC DISTURBANCE AND BOTH CR AND UR FOR THE FOUR DURATIONS OF SHOCK US

	Shock Duration (Sec.)			
	0.1	2.0	6.0	15
CR	.01	.01	.02	.01
UR	.001	.001	.01	.001

The flat gradient in Fig. 1 relating strength of CR to US duration is not inconsistent with the findings of Bitterman, Reed, and Krauskopf (1), who found no significant difference in the strength of conditioned GSR with shock duration of .5 and 3 sec. This lack of difference, however, may have appeared because they used the same *S*s for both durations. Their experiment employed different signals for short and long shocks; and, if the *S*s discriminated the signals verbally, they were assumed not to have generalized autonomically (i.e., GSR). The weakness of this assumption is shown by the fact, reported by Notterman, Schoenfeld and Bersh (5), that *S*s who state they expect no shocks during experimental extinction continue to show conditioned cardiac disturbance.

Mowrer and Solomon (4) paired a light signal with shocks of 3 sec. for one group of rats and with shocks of 10 sec. for another group. The strength of fear conditioned to the light was measured by the capacity of the light to act as effective punishment of a lever-pressing response. For these two durations of shock, an abrupt onset and termination of shock was employed. No significant difference between the two groups was found with respect to their inferred fear. Two further groups were run under a similar procedure, except that shock was terminated gradually rather than abruptly. For one group the shock was 4 sec. long and for the other, 7 sec. No reliable difference in fear was reported between these groups either. Despite the wide differences in *S*s, procedures, and methods of inferring fear, the results of the Mowrer and Solomon study are in agreement with our own findings. Such agreement gives us confidence that there is no relationship between indices of fear and noxious stimulus durations.

The discrepancy between the empirical findings and Miller's theoretical formulation of the drive-reduction view deserves comment. It will be noted that Miller prefaced his statement that short noxious stimuli would produce more fear than long ones with the phrase, "other things equal." There are at least two important "other things" here; the first is strength of US, or more precisely, *subjective* strength. This amounts to a statement of the need to control the strength of the shock, or other noxious stimulus, as it feels to the subject. If, through temporal summation, the longer shocks are felt as stronger, then stronger fear might result from them despite the correctness of Miller's statement, because more intense unconditioned stimuli are known to create stronger CRs (6, 7). The second important factor in need of control is adaptation to shock. If there is a period of rapid adaptation during the early part of a prolonged shock, it might produce the major part of drive reduction necessary to condition fear. An effect of this kind shortly after shock onset could make the length of the physical shock quite irrelevant to the timing of the drive reduction.

With respect to the first point, temporal summation, we have one way of assessing the apparent or subjective intensity of shock, that is, by the maximum or total magnitude of the unconditioned response. Our results show that there were no systematic differences in the magnitude of UR for the different shock durations. Hence, we have some evidence, at least, that the over-all subjective intensity of the US was relatively constant for the various durations. We conclude from this that the discrepancy between Miller's theoretical statement and our data can not easily be attributed to lack of control of shock intensity.

With respect to the second factor, shock adaptation, we appeal this time to the form of the unconditioned response as a relevant datum. It is true that adaptation to the shock occurs if momentary magnitude of the UR reveals subjective shock intensity. We have shown (8, 9, 10) that *regardless of shock duration* in the 0.1 – 15 sec. range the heart-rate response to shock takes the form of a rapid acceleration to a maximum rate within 4 beats of shock onset, followed by a slow irregular return (adaptation) to normal requiring more than 20 beats. If the drive property of shock parallels the course of UR magnitude in time, we would be led to say that drive reduction is remarkably unrelated to shock duration in the range investigated. In terms of Miller's application of Hull's theory, this would mean that no differences in conditioning of fear should occur because subjective shock intensity reduction (the presumed reinforcing agent) was equally delayed for all groups. If this line of reasoning is correct, our experiment unfortunately provides no real test of a theory that predicts the consequences of varying delays of reinforcement.

This difficulty, moreover, may not be restricted to the present experiment. As long as there is no demonstrable relationship between objective and subjective properties of shock, the possibility of a simple experimental test of Miller's assertion is precluded. Use of a stronger shock (if *S*s would endure long durations) might conceivably eliminate the troublesome adaptation effect, but only at the cost of a corresponding increase in the likelihood of destroying our present control of temporal summation.

The present data may be accounted for by two-factor or contiguity theories which stress the importance of the onset rather than the offset of shock (3, 4), but the data are not critical tests of this type of theory.

SUMMARY

No relationship has been found in human *S*s between conditioned heart rate response magnitude and a wide range of shock US durations.

Under the assumption that cardiac disturbance is an index of fear, this fact is related to a deduction from a drive-reduction theory of fear, according to which a signal followed by a brief noxious stimulus should require the capacity to elicit stronger fear than one followed by a prolonged noxious stimulus. It is concluded that this proposition is either incorrect or not testable under the conditions provided.

REFERENCES

1. Bitterman, M. E., Reed, P., & Krauskopf, J. The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.*, 1952, **65**, 256–262.
2. Miller, N. E. Comments on multiple-process conceptions of learning. *Psychol. Rev.*, 1951, **58**, 375–381.
3. Mowrer, O. H. Two-factor learning theory: Summary and comment. *Psychol. Rev.*, 1951, **58**, 350–354.
4. Mowrer, O. H., & Solomon, L. N. Contiguity vs. drive-reduction in conditioned fear: The proximity and abruptness of drive-reduction. *Amer. J. Psychol.*, 1954, **67**, 15–25.
5. Notterman, J. M., Schoenfeld, W. N., & Bersh, P. J. Conditioned heart rate response in human beings during experimental anxiety. *J. comp. physiol. Psychol.*, 1952, **45**, 1–8.
6. Passey, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
7. Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
8. Zeaman, D., Deane, G., & Wegner, Norma. Amplitude and latency characteristics of the conditioned heart response. *J. Psychol.*, 1954, **38**, 235–250.
9. Zeaman, D., & Wegner, Norma. The role of drive reduction in the classical conditioning of an autonomically mediated response. *J. exp. Psychol.*, 1954, **48**, 349–354.
10. Zeaman, D., & Wegner, Norma. A further test of the role of drive reduction in human cardiac conditioning. *J. Psychol.*, 1957, **43**, 125–133.

(Received November 8, 1957)

Psychological Review
Vol. 65, No. 4, 1958

STIMULUS AND RESPONSE GENERALIZATION: DEDUCTION OF THE GENERALIZATION GRADIENT FROM A TRACE MODEL[1]

ROGER N. SHEPARD

Harvard University[2]

It is now generally acknowledged that (*a*) a response conditioned to one stimulus tends also to occur to other stimuli, and (*b*) the magnitude of this response-tendency (for any particular one of those stimuli) is governed by the dissimilarity between that stimulus and the stimulus to which the response was originally conditioned. Indeed, this principle of stimulus generalization is of such fundamental importance that any quantitative theory of behavior that fails to deal with it explicitly can only be regarded as incomplete. It is not surprising, therefore, that a number of investigations have been specifically concerned with the form of the function relating generalized response-tendency to interstimulus dissimilarity. What may seem surprising, though, is that the conclusions of these various studies, far from converging on a unique function, have diverged to the several different functions illustrated in Fig. 1. Beyond those who doubt the existence of a quantitatively invariant "gradient of generalization" in the first place (**5**, **16**, **21**), there are those, like Schlosberg and Solomon, who consider this gradient to be linear, as in A (**8**, **14**, **22**, **23**); those, like Spence, who consider it to be convex upward, as in B (**27**, **28**); and those, like Hull, who consider it to be concave upward, as in C (**1**, **11**, **12**, **19**, **20**). In addition, if the "discriminal dispersion" of the psychophysicists can be interpreted as a gradient of generalization, there is support for a bell-shaped function which is first convex and then concave upward as shown in D (**9**, pp. 317–319).

In addition to stimulus generalization, an analogous phenomenon of response generalization is sometimes supposed to operate so that (*a*) a stimulus to which one response has

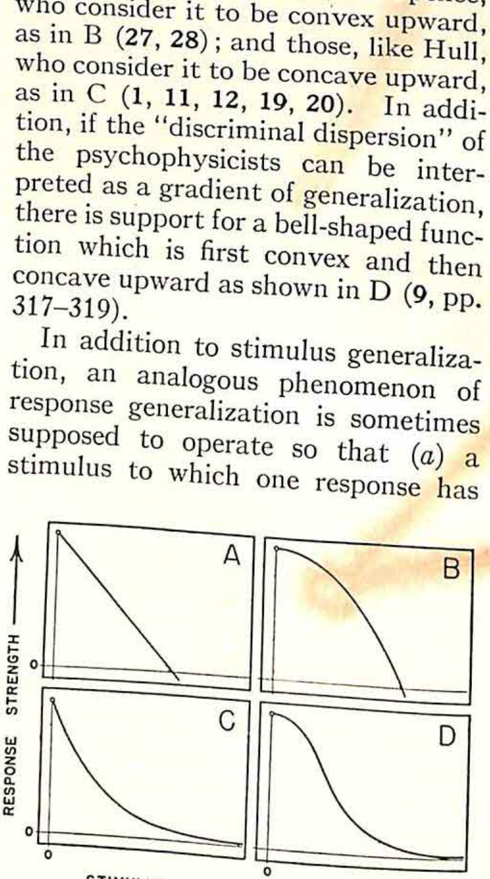


FIG. 1. Several proposals for the form of the gradient of stimulus generalization. The abscissa in each case is some measure of the dissimilarity between two stimuli, and the ordinate is some measure of the tendency of a response, conditioned to one of the two stimuli, to follow the presentation of the other.

---

[1] This work was begun on a National Academy of Sciences-National Research Council Postdoctoral Associateship at the Naval Research Laboratory, was completed under Contract AF 33(038)-14343 between Harvard University and the Operational Applications Laboratory, Air Force Cambridge Research Center, Air Research and Development Command, and appears as Rep. No. AFCRC-TN-57-62, ASTIA Document No. AD 146753. Reproduction for any purpose of the United States government is permitted. The author is indebted to H. Glaser for a number of useful suggestions and to G. A. Miller and S. S. Stevens for their improvements in the manuscript.

[2] Now at the Bell Telephone Laboratories, Murray Hill, N. J.

been conditioned tends also to evoke other responses, and (*b*) the magnitude of this tendency (for any particular one of those responses) is governed by the dissimilarity between that response and the response originally conditioned. Although this principle of response generalization is also of considerable theoretical importance, even less progress has been made toward the quantitative determination of the shape of the gradient in this case than in the case of stimulus generalization.

Actually, that a unique gradient of generalization has failed to emerge both in studies of stimulus generalization and in studies of response generalization no longer seems so surprising when one examines these studies in detail. For although the obtained gradient must depend crucially upon the choice of the independent variable (i.e., dissimilarity), the various investigators have never been able to agree on any one measure of dissimilarity as appropriate. Furthermore, since tests for generalization have been conducted during various stages of discrimination learning, as well as after various periods of extinction, one must consider the possibility that the form of the gradient may change radically under different conditions of reinforcement (**13**, **31**).

This paper will attempt to show that, by approaching the generalization problem from a somewhat different direction, considerable evidence can be adduced for the proposition that the gradients of stimulus and response generalization both conform to an exponential decay function (Curve C of Fig. 1). Further evidence will be presented to indicate that the form of the gradient does indeed depend upon the schedule of reinforcement and, more particularly, that it changes from the exponential to a bell-shaped function (Curve D) as nonreinforced trials are introduced with greater and greater frequency.

THE EXPONENTIAL GRADIENT IN PAIRED-ASSOCIATE LEARNING

In view of the difficulties attending any effort to establish one measure of dissimilarity as a standard, the following strategy was recently proposed: Stimuli (or responses) were conceptualized as points in a "psychological space" in such a way that the distance between any pair of these points represented the psychological dissimilarity between the corresponding pair of stimuli (or responses). By taking account of a set of metric axioms which any measure of distance should satisfy, it was shown that hypotheses about the shape of the gradient of generalization could be tested without resorting to an independent measure of dissimilarity (**25**). In a series of experiments on stimulus and response generalization during paired-associate learning, substantial support was obtained for the hypothesis that generalization is an exponential decay function of psychological distance (**26**). Since alternative hypotheses were not tested, however, it seems desirable to present data in a form that will clearly rule out the Functions A, B, and D of Fig. 1.

Now in a paired-associate experiment there are N stimuli, $S_1, S_2, \ldots S_N$, to which are assigned N responses, $R_1, R_2, \ldots R_N$. On each trial one of the N stimuli is presented to the subject who must in turn produce one of the N responses. Various procedures of differential reinforcement can be used to communicate to the subject the prevailing assignment of the responses to the stimuli. Sometimes the subject is simply informed after each trial whether his response was or was not the correct one (i.e., the one assigned to the stimulus presented on that trial); at other times more elaborate methods of "correction" are used (26). In any case the so-called assignment is simply a rule which the experimenter follows in delivering the reinforcements, and any arbitrary one-to-one assignment can be established in this way.

Consider then a paired-associate experiment in which the responses are highly distinctive and so lead to negligible amounts of generalization. If there are N pairs consisting of one stimulus and its assigned response, the data from such an experiment can be represented by an $N \times N$ matrix giving, for every S_i and S_k, the conditional probability $P_{ik}{}^S$ with which S_i leads to the response assigned to S_k (25). A very basic measure of stimulus generalization between S_i and S_k is then provided by either of the probabilities $P_{ik}{}^S$ or $P_{ki}{}^S$. Comparison between different experiments is facilitated, however, by adjusting these measures so that the generalization between any stimulus and itself is always unity. This can be done by replacing the absolute probabilities $P_{ik}{}^S$ and $P_{ki}{}^S$ with the ratios $P_{ik}{}^S/P_{ii}{}^S$ and $P_{ki}{}^S/P_{kk}{}^S$. Furthermore, these two ratios can be averaged together to furnish a single, more stable estimate of the generalization between S_i and S_k. Since there are theoretical reasons for preferring the geometric to the arithmetic mean (25), the measure of stimulus generalization might be defined by the formula

$$\sqrt{\frac{P_{ik}{}^S P_{ki}{}^S}{P_{ii}{}^S P_{kk}{}^S}}$$

Owing to the gradual manner in which differential reinforcement takes effect, however, there is an initial phase during which the subject responds more or less randomly. This means that the function given in the above formula necessarily levels off at some positive asymptote for large interstimulus distances. Again, in order to compare data from different experiments, this asymptote must be brought down to zero for each experiment. This is accomplished by estimating a parameter C^S from each set of data (25), and by defining the generalization between S_i and S_k to be

$$G_{ik}{}^S = (1 + C^S) \times \sqrt{\frac{P_{ik}{}^S P_{ki}{}^S}{P_{ii}{}^S P_{kk}{}^S}} - C^S \quad [1]$$

Likewise, in an experiment with highly discriminable stimuli, the response generalization between R_i and R_k will be given by

$$G_{ik}{}^R = (1 + C^R) \times \sqrt{\frac{P_{ik}{}^R P_{ki}{}^R}{P_{ii}{}^R P_{kk}{}^R}} - C^R \quad [2]$$

where $P_{ik}{}^R$ is the conditional probability of R_k, given the stimulus assigned to R_i (25).

Equations 1 and 2 then specify the dependent variables for paired-associate experiments, i.e., stimulus and response generalization. With regard to the independent variables, i.e., interstimulus and interresponse distances, it is clear that physical measures are not directly applicable. Thus the psychological distance between two tones at a fixed difference in intensity changes as both tones are increased in intensity. But this does not mean that there is no relation between physical and psychological distance. On the contrary, at least the following two statements can be made: (*a*) two tones which are brought arbitrarily close together in terms of physical measures also approach each other psychologically; (*b*) as two tones separated by a fixed difference in intensity are increased in intensity, the psychological distance (although it does not remain fixed) nevertheless changes in a gradual and continuous manner. These considerations form a basis for the assumption that the locus of a set of stimuli in psychological space can always be obtained from their locus in physical space by a transformation that is (*a*) continuous and (*b*) differentiable (25).

Although the stimuli and responses chosen for experiments on paired-associate learning have typically been words or nonsense syllables, nonverbal materials can be used just as well. Indeed, for the purposes of establishing a measure of the psychological distance between stimuli, it is especially convenient to choose stimuli (such as tones differing only in intensity) that can be varied along a single physical dimension. In general, of course, stimuli that are evenly spaced along a single physical dimension will be neither evenly spaced nor confined to a straight line in psychological space. Nevertheless, it follows from the assumptions of continuity and differentiability that, if stimuli are evenly spaced over a sufficiently small range of a single physical dimension, then they will be spaced approximately evenly along an approximately straight line in psychological space.

Suppose, then, that the N stimuli of a paired-associate experiment satisfy this special condition which, for present purposes, will be called the linearity condition. Such stimuli can be designated as $S_1, S_2, \ldots S_N$ in such a way that the subscripts correspond to the ordinal positions of the N stimuli along the common (approximately straight) line. The average generalization for all pairs of stimuli just D steps apart along this line will then be given by

$$G(D) = \frac{1}{N - D} \sum G_{ik} \qquad (\text{with } i - k = D) \quad [3]$$

where, as indicated, the summation is carried out over the $N - D$ stimulus pairs separated by just D steps. Similarly, if the responses meet the linearity condition, this formula can be used as a measure of the average generalization for all pairs of responses separated by D steps. Thus, G_{ik} in Formula 3 can stand for either $G_{ik}{}^S$ or $G_{ik}{}^R$.

Now the average generalization $G(D)$ and the separation D have properties that particularly recommend them for investigations of the relation between generalization and distance. First of all, D is quite insensitive to the physical measure used to space the stimuli evenly along the chosen dimension. Thus, if the stimuli are squares differing only in size, it does not much matter whether these squares are spaced in accordance with constant increments of area or length of side. It is a consequence of the linearity condition that these two variables (a^2 and a) will be almost linearly related within the small range of variation permitted along the size dimension. In fact, any variable that is equivalent to these variables except for a continuous, differentiable transformation could presumably be used. Thus the stimuli could just as well be regularly spaced in terms of measures based upon psychophysical procedures such as the summation of jnd's, category judgment, magnitude estimation, etc. (29). All of these measures are continuous and differentiable functions of the usual physical variables.

This indifference of D to the variable chosen for evenly spacing the stimuli is further enhanced by com-

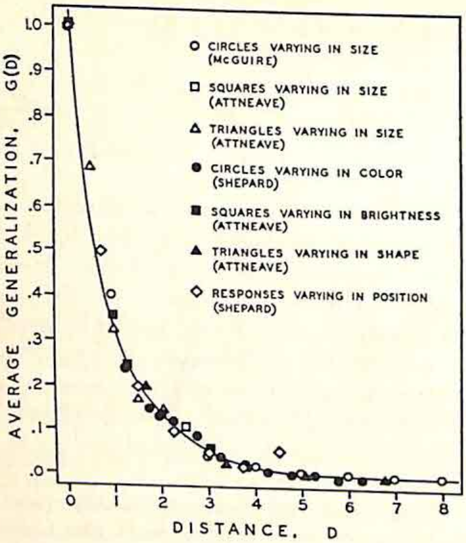


FIG. 2. An exponential decay function fitted to data from several experiments involving stimulus or response generalization during paired-associate learning. The center of each plotted symbol (triangle, square, circle) has average generalization $G(D)$ as ordinate and distance D (multiplied by a constant factor) as abscissa. The experiments were conducted by Attneave (1), McGuire (18, 26), and Shepard (26). In the experiment with circles of variable color, the stimuli were not confined to a single physical dimension. Since the linearity condition was not met in this case, the D-values were taken from a multidimensional scaling solution obtained by Torgerson for these same colors (30, Ch. 11). Since the stimuli were contained in a relatively small region of psychological space, it seems safe to assume that Torgerson's judgmental method of triads yields satisfactory estimates of psychological distance in the present sense. In the experiment on response generalization, it had been found that the two end-responses did not conform to the linearity requirement (26). The plotted data were therefore based upon the generalization between the seven intermediate responses only. Finally, Attneave's experiments deviated from the design presupposed by the present analysis in the following respects. (*a*) The stimuli were quite widely spaced and so may not have met the linearity requirement. (*b*) The number of trials was much smaller than in the other experiments. This makes the steady-state condition assumed in the ensuing theoretical discussion seem a little unlikely. (*c*) Since Attneave reported only the sums $P_{ik}{}^S + P_{ki}{}^S$ but not the individual probabilities, it was necessary to substitute the approximate formula $(P_{ik}{}^S + P_{ki}{}^S)/(P_{ii}{}^S + P_{kk}{}^S)$ for the geometric mean used in Equation 1. However, in a previous investigation this approximation was very close and, indeed, possessed greater statistical stability than the geometric mean (24). (*d*) There were not sufficient data to make the estimation of the constants C^S feasible. For the purposes of plotting the data, therefore, C^S was assumed to be zero. That the agreement between the various sets of data is good despite the deviations from optimum conditions in certain instances suggests that the data may not be particularly sensitive to such deviation.

puting $G(D)$ as an average for all pairs of stimuli separated by D steps. Thus, even if there is some residual systematic contraction or expansion of psychological distance, as one proceeds from the first to the last pair of stimuli at a given separation D, this systematic effect will be largely canceled out by averaging both ends of the stimulus range together.

In Fig. 2 average generalization $G(D)$ is plotted against D for several experiments on the learning of paired associates. In each experiment several different random assignments of the responses to the stimuli were employed. Since the average psychological spacing of the stimuli and responses varies somewhat from experiment to experiment, the D-values for each experiment are multiplied by a constant factor to make them comparable with the D-values for the other experiments. The conclusion is clear: Although the empirical points fall closely along an exponential decay function (the fitted curve), they deviate markedly from the alternative Functions A, B, and D of Fig. 1.

DEDUCTION OF THE EXPONENTIAL GRADIENT FROM A TRACE MODEL

The principal purpose of this paper is to propose a model to account for the empirically determined exponential form of the gradient of generalization in terms of a hypothetical trace process. The aim of such a proposal is twofold: First, it is desired to remove the apparent arbitrariness of the exponential gradient by showing that it follows from certain elementary assumptions of a more intuitively compelling character. Second, it is hoped that such a model will provide for prediction to other experiments in which the usual paired-associate conditions no longer prevail.

The model is suggested by recognition experiments of the following kind. A subject is shown a certain square and, after a delay of t units of time, shown a second square which may or may not be the same as the first. In this situation the probability of responding "same" is distributed, with respect to the difference in size between the first and second squares, according to some bell-shaped density function. Since the time error is usually small, the mode falls near the zero-difference point. The variance of the distribution, however, increases appreciably with the delay t imposed between the first and second exposure (2, 3, 10, 17).

The unidimensional stimuli, S_1, S_2, . . . S_N, in an experiment on paired-associate learning, will be designated by small circles arranged in a vertical row along the left, as in A of Fig. 3. Corresponding to these, there are conceived the internal representations (perceptions), S_1^*, S_2^*, . . . S_N^*, which will be designated by the small circles on the right. Whenever a stimulus S_i is presented, it leads to some one of the internal representations S_k^*, with probability P_{ik}. If the responses are so distinctive as never to be confused, reinforcement will ensure only if a stimulus is followed by its corresponding representation (25).

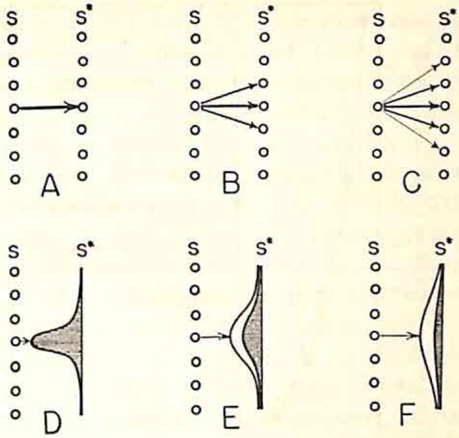


FIG. 3. The diffusion and deconditioning of the stimulus trace. The arrows connecting a single stimulus S to the alternative perceptual representations S^* in A, B, and C illustrate how the trace elements are assumed to spread out with time. In D, E, and F the process is represented as continuous.

Suppose, then, that when S_i leads to S_i^*, a large number of trace elements are conditioned (by reinforcement) from S_i to S_i^*. This bundle of trace elements or "stimulus trace" is designated by the arrow in A. Immediately following the removal of the external stimulus S_i, the trace elements are subject to haphazard perturbations so that some of the conditioned trace elements wander to adjacent stimulus representations as shown in B. (One might imagine here the action on the synapses of the random molecular processes associated with metabolism.) Later still, some of these elements will wander even further from S_i^*, as shown in C.

Now for any two stimuli differing along a single physical dimension, another stimulus can be found which is situated intermediately between these. Thus, as indicated in D, a continuum may be conceived as underlying the corresponding internal representations. This continuum then is the psychological space of the stimuli.

After removal of the stimulus, the distribution of trace elements progressively spreads out in this space, as illustrated in D, E, and F.

In addition to this spread or diffusion of the trace, it is assumed that the trace elements are also subject to spontaneous deconditioning (again, presumably owing to random processes at the molecular level). Thus, as the elapsed time increases, the number of elements still conditioned decays to zero. In Fig. 3, the shaded areas represent the fraction of the original trace elements which are still conditioned at each time.

The Deconditioning of the Trace

The probability that a given conditioned element will suffer deconditioning during a small interval Δt will be denoted by $U(\Delta t)$. The simplest rule that can be assumed to govern this probability is as follows:

Assumption I. $U(\Delta t)$ is a constant, independent of the time the element has remained conditioned and independent of the distance (in psychological space) to which the element has drifted in that time.

The probability that an element, still conditioned at time t, remains conditioned until $t + \Delta t$ is, of course, $1 - U(\Delta t)$. Therefore, the probability that a given trace element remains conditioned during the first m intervals of length Δt, but then becomes deconditioned during the immediately succeeding Δt-interval, is, by Assumption I,

$$[1 - U(\Delta t)]^m U(\Delta t) \qquad [4]$$

Now $U(\Delta t)$ is necessarily proportional to the interval chosen for Δt. It is convenient, therefore, to define a deconditioning parameter,

$$U = \lim_{\Delta t \to 0} \frac{U(\Delta t)}{\Delta t} \qquad [5]$$

which does not depend upon the choice of an arbitrary interval Δt. Then, if the t-axis is translated so that conditioning takes place at $t = 0$, m can be increased in such a way that, as $\Delta t \to 0$, $m \cdot \Delta t \to t$.

Although the probability that a given trace element is deconditioned at precisely time t is zero, the probability per unit time (the "probability density") or rate of deconditioning for an individual trace element is, at the particular instant t,

$$\lim_{\Delta t \to 0} [1 - U(\Delta t)]^{t/\Delta t} U(\Delta t)/\Delta t$$

But by Equation 5 as $\Delta t \to 0$, $U(\Delta t) \to U \cdot \Delta t \to Ut/m$. Therefore, the rate of deconditioning at time t is, for single trace elements,

$$\lim_{m \to \infty} [1 - Ut/m]^m U$$

$$= U \exp(-Ut) \qquad [6]$$

where, for convenience in what follows, $\exp(-Ut)$ is used in place of e^{-Ut}.

With regard to the deconditioned trace elements, the following rule is the simplest that can be assumed:

Assumption II. In the absence of further reinforcements, a deconditioned element remains deconditioned.

From this assumption and Equation 6 it follows that the fraction of the originally conditioned elements remaining conditioned at time t is

$$U(t) = 1 - \int_0^t U \exp(-U\tau)d\tau$$

$$= \exp(-Ut) \qquad [7]$$

The Diffusion of the Trace

Equation 7 completes the quantitative formulation of the deconditioning of the trace. In order to provide a similar formulation for the diffusion of the trace, it is necessary to consider the motions of the individual trace elements in psychological space. The exposition is simplified by continuing to suppose that the stimuli are evenly spaced along a restricted range of a single physical dimension. It is then possible to introduce one coordinate x_i for each stimulus representation S_i^* giving its position along a one-dimensional psychological space. The psychological distance between any two stimuli, S_i and S_k, is then

$$D_{ik} = |x_i - x_k| \qquad [8]$$

Now the expression $V_{ik}(\Delta x, \Delta t)$ will denote the probability that, if at time t a trace element is situated at x_i, by time $t + \Delta t$ it will have moved into the one-dimensional region bounded by x_k and $x_k + \Delta x$. As before, the arbitrary interval Δx can be eliminated by defining a new quantity

$$V_{ik}(\Delta t) = \lim_{\Delta x \to 0} \frac{V_{ik}(\Delta x, \Delta t)}{\Delta x} \qquad [9]$$

$V_{ik}(\Delta t)$ is the probability density of displacements from x_i to x_k during the brief interval Δt. The simplest rule that can reasonably be assumed to govern this quantity is as follows:

Assumption III. $V_{ik}(\Delta t)$ is an invariant function of the psychological distance between S_i and S_k. It does not depend upon the time t or upon the absolute position of the pair, S_i and S_k, in psychological space.

If the x-axis is translated so that $x_i = 0$, it follows from Assumption III that, for a given interval Δt, there exists a fixed function $f_{\Delta t}$, such that

$$V_{ik}(\Delta t) = f_{\Delta t}(D_{ik}) = f_{\Delta t}|x_k| \qquad [10]$$

However, it will not be necessary to make any particular assumption about the form of the function $f_{\Delta t}$. It is only necessary to insure that, during a short period of time, the probability of a very large displacement is negligible. This can be stated with greater precision as follows:

Assumption IV. The function $f_{\Delta t}$ and, thus, the probability density of displacements of a trace element from $x = 0$ is distributed over the x-axis with finite variance.

From Assumptions III and IV it follows that the variance of the distribution of displacements during an interval Δt is the finite constant

$$V(\Delta t) = \int_{-\infty}^{+\infty} x_k^2 V_{ik}(\Delta t) dx_k \qquad [11]$$

Since $V(\Delta t)$ must depend upon the length of the interval Δt, it is useful to define a diffusion parameter,

$$V = \lim_{\Delta t \to 0} \frac{V(\Delta t)}{\Delta t} \qquad [12]$$

which does not require the stipulation of an interval Δt. Just as U governs the rate of deconditioning of the trace, then, V controls the rate of spread or diffusion of the trace in psychological space.

The next question to be answered is this: Given the diffusion parameter V, what form will the distribution of trace elements take after an appreciable delay t? It is possible to show that the assumptions which have been made are sufficient conditions for the desired distribution to tend toward a limiting form which is independent of the form of $f_{\Delta t}$ (6, 15). Specifically, if a trace comprising a large number (n) of elements is conditioned from S_i to x_i at $t = 0$, the density of these elements at x_k for some later time t conforms with the Gaussian function

$$n \cdot V_{ik}(t) = n \cdot (2\pi V t)^{-\frac{1}{2}} \times \exp\left[-(D_{ik})^2/2Vt\right] \qquad [13]$$

The beauty of this result is its complete independence from the underlying mechanism symbolized by $f_{\Delta t}$. Thus, even though one imagines, for example, that thermal agitation of the molecular substrate is responsible for deconditioning and diffusion, the biophysical details of this process need not be specified. For, according to the present formulation, these details are irrelevant to the question of the gross behavior of the trace system.

The Trace Process in Paired-Associate Learning

Now in the course of learning paired associates, each stimulus S_i will have been presented on many occasions. Furthermore, on a number of these occasions, reinforcement of the response assigned to S_i will have conditioned a bundle of n trace elements to S_i^*. Thus at some given time t_0, the density of conditioned elements resulting from the immediately preceding reinforcement will be distributed in psychological space as shown for t_{-1} in Fig. 4. Likewise, the densities remaining from earlier reinforcements will be distributed as shown for t_{-2}, t_{-3}, and so on.

The total distribution of conditioned elements emanating from S_i at t_0 can be found by summing the Gaussian distributions resulting from all previous reinforcements of the response assigned to S_i. It is possible to derive an analytic approximation to this composite distribution ($\int_t$ in Fig. 4) by supposing that, after an initial phase of rapid learning, reinforcements occur at relatively frequent and regular intervals. If the early trials are disregarded, then, a roughly steady-state process can be considered. The summation of the Gaussian curves of various ages can then be approximated by an integration of these curves over t.

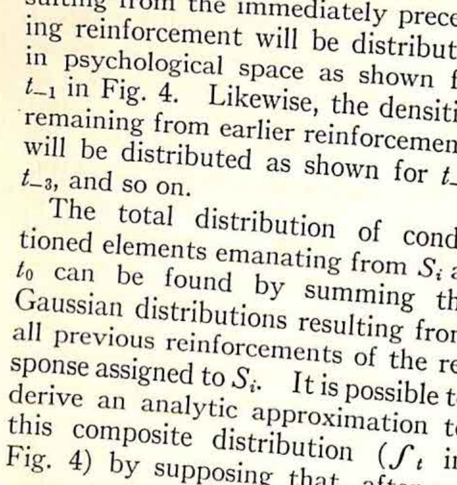


FIG. 4. A series of Gaussian distributions with increasing variances, and the composite distribution arising from an integration of these over time.

Now, with respect to S_i, the density of trace elements at x_k resulting from a reinforcement t time units ago is given by $n\cdot V_{ik}(t)$. However, only the fraction $U(t)$ of these is still conditioned. Therefore the density of elements at x_k which remain conditioned after a delay t from reinforcement is $n\cdot U(t)\cdot V_{ik}(t)$. Clearly, then, the total density of conditioned trace elements resulting from all previous reinforcements of the response assigned to S_i is distributed in approximate accord with

$$\int_0^\infty n\cdot U(t)\cdot V_{ik}(t)dt$$

$$= \int_0^\infty n\cdot\exp\,(-\,Ut)\cdot(2\pi Vt)^{-\frac{1}{2}}\cdot\exp\,[-\,(D_{ik})^2/2Vt]dt \quad [14]$$

Fortunately, the integration can be effected (**4**, p. 144) and, indeed, yields

$$n(2UV)^{-\frac{1}{2}}\cdot\exp\,[-\,D_{ik}(2U/V)^{\frac{1}{2}}] \quad [15]$$

For intermediate stages of paired-associate learning, then, this function can be taken as a measure of the strength of connection of the stimulus S_i to the point x_k on the perceptual continuum. It is not a probability density, however, since (if $n > 1$ or $U < 1$) the integration of this function over the x-axis yields a value greater than unity (**7**, p. 133). Multiplication of Equation 15 by U/n converts this function to a probability density,

$$P_{ik} = (U/2V)^{\frac{1}{2}} \cdot \exp\left[- D_{ik}(2U/V)^{\frac{1}{2}}\right] \quad [16]$$

for then

$$\int_{-\infty}^{+\infty} P_{ik} dx_k = 1 \quad [17]$$

Now Equation 16 furnishes an estimate of the conditional probability (per unit x) of the particular perception x_k, given the stimulus S_i. In order to secure the probability of taking S_i to be S_k (through stimulus generalization), $S_k{}^*$ can be reinterpreted as a finite region partitioned off from psychological space in the neighborhood of x_k. In this way the entire one-dimensional space can be divided into N mutually exclusive and exhaustive segments so that each segment corresponds to one of the external stimuli. A given trace element is then said to be conditioned to $S_i{}^*$ at time t if and only if it falls in the region containing x_i at that time.

P_{ik}, reinterpreted in this way, can be taken as an approximation to the conditional probability that the external stimulus S_i will lead to the internal representation $S_k{}^*$. If the early trials are ignored (since these trials do not sufficiently approach a steady state), the constant C^S in Equation 1 can be disregarded (25). Setting $C^S = 0$, and substituting the right-hand member of Equation 16 for $P_{ik}{}^S$ in Equation 1, then, the generalization between S_i and S_k assumes the remarkably simple form

$$G_{ik}{}^S = \exp\left(- D_{ik} \sqrt{2U/V}\right) \quad [18]$$

Letting κ stand for the constant $\sqrt{2U/V}$ and averaging $G_{ik}{}^S$ over all pairs of stimuli separated by a fixed distance D, Equation 3 now takes the form

$$G(D) = e^{-\kappa D} \quad [19]$$

But if κ is identified with the constant distance multiplier calculated for each experiment, this is precisely the exponential function fitted to the empirical data in Fig. 2.

Further Aspects of the Trace Process

The role of deconditioning. It might have seemed unnecessary to include the deconditioning assumptions (I and II) along with the diffusion assumptions (III and IV), since the process of diffusion alone would account for the spread of the trace, and hence for stimulus generalization. However, from Equation 18 it is clear that, without deconditioning, generalization would be so extensive as to prohibit the learning of paired associates. For if $U = 0$, $G_{ik} = 1$ for all i and k. In this case the gradient is perfectly flat, and discrimination between stimuli is impossible. This is a consequence of carrying the integration of Equation 14 out to infinite t. Alternatively, one could integrate only out to some finite value $t = T$. However, in terms of the model, this is equivalent to assuming that the independent elements simultaneously suffer deconditioning at the same instant T time units from conditioning. Rather than postulate coincidences of this kind, it seems more plausible to assume the gradual kind of fading away implied by Assumptions I and II. This fading away (or forgetting) then serves the adaptive function of weighting old, diffuse traces less heavily than new, accurate traces.

Of course, the integration of Equation 14 out to infinite t is not strictly justified, since the learning experiment itself proceeds for only a finite period. The error introduced by the infinite integration is small, however, if the deconditioning parameter U is sufficiently greater than zero. For then the hypothetical traces persisting from reinforced trials that might be imagined as preceding the actual beginning of the experiment would be almost totally deconditioned during all but the early trials of the learning experiment.

Asymmetric generalization. It has sometimes been suggested that there exist asymmetries in which a generalization going in one direction, e.g., $S_i \rightarrow S_k^*$, is more probable than a generalization going in the reverse direction, $S_k \rightarrow S_i^*$. At first glance, such a possibility appears to violate Assumption III, according to which the probability of a displacement for a given trace element depends only upon the length (and not upon the direction) of that displacement. However, an account of such asymmetries which is consistent with Assumption III is suggested by the analysis proposed by Bush and Mosteller. According to their model, the psychological dissimilarity going from S_i to S_k is equal to the dissimilarity going from S_k to S_i only if the "set of stimulus elements" comprising S_i has the same measure as the set of stimulus elements comprising S_k (5). In terms of the trace model, then, it can be supposed that the area of the region of psychological space designated as S_i^* may be greater or smaller than the area of the region designated as S_k^*. Since the probability that a given trace element (wandering at random) will occupy a certain region during an interval Δt is proportional to the area of that region, P_{ik}^S does not necessarily equal P_{ki}^S. Indeed, the weight W_i^S, proposed earlier by Shepard (25), is presumably a measure of the area of the region corresponding to S_i. This interpretation suggests why the empirically obtained weights tended to be greatest for stimuli or responses at the end of a linear array (26).

Multidimensional generalization. For expository reasons the stimuli were supposed to vary along a restricted range of a single physical dimension. Actually, by making use of the theory of random motions in Euclidean spaces of more than one dimension (6), it can be shown that the trace model leads to the same exponential gradient in either of the two following multidimensional cases: (a) the psychological space is Euclidean; (b) the stimuli are confined to a small region of psychological space. (The second case follows from the first. For, by the hypothesized relation between physical and psychological space, a sufficiently small region of even a non-Euclidean space will be approximately Euclidean.) The treatment of generalization over large distances in non-Euclidean spaces, however, awaits the development of a general theory of random motions in spaces of positive and negative curvatures.

Response generalization. The preceding discussion has been formulated solely in terms of stimulus generalization. However, the same trace model can also be applied in the case of response generalization. Suppose, for example, that the responses are closely spaced along a single physical dimension, whereas the stimuli are completely discriminable. When one response R_i is intended, some other response R_k may actually be made. Using a notation analogous to that adopted for stimulus generalization, it can be said that R_i^* leads to R_k with probability P_{ik}^R. If then R_i^* leads to R_i, the ensuing reinforcement conditions a large number of trace elements from R_i^* to a point x_i in the psychological space of the responses. These conditioned trace elements are then subject to the rules already set forth in Assumptions I through IV.

THE FORM OF THE GRADIENT AND THE SCHEDULE OF REINFORCEMENT

In the last section the apparent arbitrariness of the specific function fitted to the paired-associate data of Fig. 2 was reduced by showing that this function is a consequence of elementary assumptions which do not involve the specification of any particular function. The purpose of this section is to determine whether these same assumptions have any consequences for different types of experiments in which the sequence of reinforced and nonreinforced trials is manipulated, as in the study by Humphreys (**13**).

Certainly, from Fig. 4 it is clear that the sharp peak of the combined gradient at $x = 0$ is contributed solely by the distributions of trace elements resulting from the most recent reinforcements (like the distribution t_{-1}). Therefore, if feedback as to the correctness of each response is terminated, the composite gradient should become rounded (convex upward) in the vicinity of $x = 0$ and, under continued extinction, gradually flatten out. Likewise, if reinforcement is delivered only for very nth correct response, the composite gradient should assume an intermediate form which tends more toward either an exponential or a bell-shaped curve as n is made smaller or larger. Considerations of this kind may account for the finding of Humphreys that, whereas with 100 per cent reinforcement the generalization gradient was concave upward, with 50 per cent reinforcement it was initially convex upward (**13**).

It is also feasible to analyze data from paired-associate experiments with 100 per cent reinforcement for this effect. Since the stimuli are presented in random order, the delay between successive occurrences of a given stimulus or response varies over a considerable range. Thus separate gradients can be plotted depending on the number of trials intervening between a given stimulus-response sequence and the most recent feedback concerning the correctness of that particular sequence. The most extensive data currently available in a form suitable for this analysis come from an experiment on response generalization in paired-associate learning (**26**). Those data are therefore reanalyzed and plotted in Fig. 5. As predicted from the model, the gradient of generalization systematically changes from concave to convex upward in the vicinity of the correct response as the number of trials intervening between a response and the

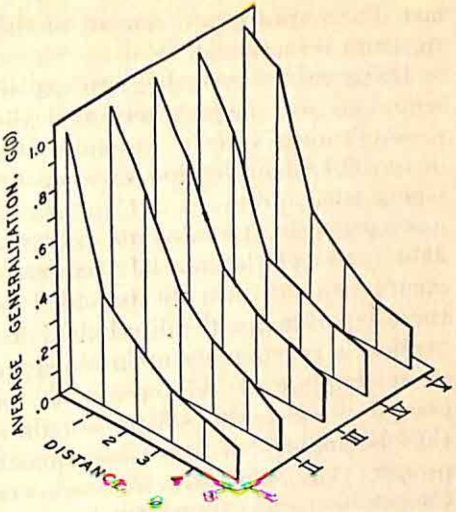


FIG. 5. The gradient of generalization as a function of the number of trials intervening between a response and the last preceding feedback as to the correctness of that response. The numbers of intervening trials are grouped together as follows: I. 0 trials; II. 1–3 trials; III. 4–8 trials; IV. 9–15 trials; and V. 16 or more trials.

last differential reinforcement of that response is increased.

These results may help to explain why the so-called "discriminal dispersion" observed in absolute-judgment and identification experiments seems to conform to a Gaussian or normal density function (**9**, pp. 317–319). Since differential reinforcements are not usually provided in these experiments, the discriminal dispersion is presumably maintained by those haphazard reinforcements of everyday existence which antedate the beginning of the experiment proper. Under these circumstances the gradient must be relatively mesokurtic, as illustrated in V of Fig. 5. This gradient resembles a Gaussian function, and is quite unlike the comparatively leptokurtic gradients (such as I) which have been shaped by recent differential reinforcement.

MICROMECHANICAL AND MACROMECHANICAL MODELS FOR GENERALIZATION

The present trace model might be termed a *micromechanical model* because it is based upon assumptions about the fine-grain or "microscopic" mechanics of the underlying trace process. This model is to be distinguished from an earlier *macromechanical model* derived from assumptions pertaining to large-scale or "macroscopic" aspects of generalization (**25**). However, to the extent that the two models are compatible, they can be integrated so that the macromechanical assumptions of the earlier model will appear as deductions from the more primitive micromechanical assumptions of the present model. The aim of such an integration would be to broaden the scope of the earlier model. For, whereas the macromechanical model applies only to paired-associate experiments with continual differential reinforcement, the micromechanical model has implications for a considerably wider range of experiments.

The three basic assumptions from which the macromechanical model was originally derived are as follows. (*a*) Stimulus and response generalizations take place independently of each other. (*b*) The probability of a stimulus generalization is an exponential decay function of the psychological distance between the stimuli. (*c*) The probability of a response generalization is an exponential decay function of the psychological distance between the responses. Now to say that the scope of the earlier model will be increased by the adjunction of the micromechanical assumptions is to say that the macromechanical assumptions (*a*, *b*, and *c*) will retain their original form only under certain limiting conditions (e.g., under continual reinforcement). When these conditions are modified (as when the reinforcements are delivered only intermittently), Propositions *b* and *c* will have to assume different forms in accordance with the conclusions of the last section.

Even with continual reinforcement, there is one case of paired-associate learning for which the macromechanical assumptions would have a different form if deduced from the micromechanical assumptions. Specifically, if the interstimulus distances and the interresponse distances are both quite small, the trace model does not lead directly to the exponential gradient. Because occasional sequences of the form $S_i \rightarrow S_k^* \rightarrow R_k^* \rightarrow R_i$ will be reinforced, trace elements will be conditioned from S_i to x_k (rather than only to x_i). Events of this kind will somewhat alter the form of the gradient. (In a previous experiment with generalization, both between the stimuli and between the responses, no significant departure from prediction on the assumption of an exponential gradient was observed [26]. However, the theoretically expected deviations would be quite small and may have been obscured by the rather large variability of the data from that experiment.)

Detailed derivations from the micromechanical assumptions are quite complicated in the case of simultaneous stimulus and response generalization, owing to the circumstance that, in this case, the form of the gradient depends upon the particular stimulus-response assignment enforced. Part of this complication is connected with the absence in both models of any account of the decrease in generalization which necessarily accompanies learning. An entirely satisfactory treatment of these problems will probably require an even more basic integration of these models with the already extensively developed models for learning per se.

SUMMARY

The problem of the relation between generalization and dissimilarity (i.e., the problem of the shape of the "gradient of generalization") is reexamined in the light of recent theoretical and empirical developments. With regard to experimental arrangements in which reinforcments are delivered in accordance with a one-to-one assignment of the responses to the stimuli (as in paired-associate learning), the following conclusions are drawn:

1. Measures of generalization can be defined in terms of the conditional probabilities with which the various stimuli lead to the various responses.
2. Thus defined, stimulus generalization and response generalization are both invariant functions of interstimulus and interresponse dissimilarities, respectively, provided that two conditions are met. First, dissimilarity is reinterpreted to mean a "psychological distance" which (*a*) is equivalent to "physical distance" except for a continuous, differentiable transformation, and (*b*) satisfies the metric axioms. Second, a given schedule of reinforcement is maintained.
3. Under conditions of frequent and regular reinforcement (as in the typical paired-associate experiment), the gradient of generalization is closely approximated by an exponential decay function (concave upward).
4. Under conditions of infrequent or intermittent reinforcement, this gradient departs from the exponential function in that it is convex upward in the vicinity of the reinforced stimulus or response.
5. The empirically observed gradients of generalization can be deduced from a mathematical model based upon four elementary assumptions concerning the temporal decay of stimulus and response traces.

REFERENCES

1. ATTNEAVE, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, 63, 516–556.
2. BACHEM, A. Time factors in relative and absolute pitch determination. *J. acoust. Soc. Amer.*, 1954, 26, 751–753.
3. BALDWIN, J. M., & SHAW, W. J. Memory for square-size. *Psychol. Rev.*, 1895, 2, 236–239.
4. BIERENS DE HAAN, D. *Nouvelles tables d'intégrales définies.* New York: G. E. Stechert, 1939.
5. BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, 58, 413–423.
6. CHANDRASEKHAR, S. Stochastic problems in physics and astronomy. In N. Wax (Ed.), *Noise and stochastic processes.* New York: Dover, 1954. Pp. 3–89.

7. FELLER, W. *An introduction to probability theory and its applications.* New York: Wiley, 1950.
8. GRANDINE, LOIS, & HARLOW, H. F. Generalization of the characteristics of a single learned stimulus by monkeys. *J. comp. physiol. Psychol.*, 1948, **41**, 327–338.
9. GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.
10. HARRIS, J. D. The decline of pitch discrimination with time. *J. exp. Psychol.*, 1952, **43**, 96–99.
11. HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
12. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
13. HUMPHREYS, L. G. Generalization as a function of method of reinforcements. *J. exp. Psychol.*, 1939, **25**, 361–372.
14. JOHNSON, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34**, 425–436.
15. KAC, M. Random walk and the theory of Brownian motion. In N. Wax (Ed.), *Noise and stochastic processes.* New York: Dover, 1954. Pp. 369–391.
16. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
17. LEYZOREK, M. Two-point discrimination in visual space as a function of the temporal interval between the stimuli. *J. exp. Psychol.*, 1951, **41**, 364–375.
18. McGUIRE, W. J. A multi-process model for paired-associates learning. Unpublished doctoral dissertation, Yale Univer., 1954.
19. MARGOLIUS, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.
20. PLOTKIN, L. Stimulus generalization in Morse code learning. *Arch. Psychol.*, 1943, **40**, No. 287.
21. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
22. ROTHKOPF, E. Z. A measure of stimulus similarity and errors in some paired-associate learning tasks. *J. exp. Psychol.*, 1957, **53**, 94–101.
23. SCHLOSBERG, H., & SOLOMON, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33**, 22–39.
24. SHEPARD, R. N. Stimulus and response generalization during paired-associates learning. Unpublished doctoral dissertation, Yale Univer., 1955.
25. SHEPARD, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.
26. SHEPARD, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509–523.
27. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
28. SPENCE, K. W. A reply to Dr. Razran on the transposition of response in discrimination experiments. *Psychol. Rev.*, 1939, **46**, 88–91.
29. STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.
30. TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.
31. WICKENS, D. D., SCHRODER, H. M., & SNIDE, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–56.

(Received December 17, 1957)

VOL. 65, NO. 5 SEPTEMBER, 1958

THE PSYCHOLOGICAL REVIEW

THE EFFECTS OF EARLY HANDLING ON VIABILITY OF THE ALBINO RAT[1]

EVERETT W. BOVARD

Montreal Neurological Institute

This paper concerns effects of early handling on viability of the albino rat. The relation of early experience, particularly handling, to the response to stress will be considered first; then neurophysiological mechanisms mediating the observed effects of handling on viability of the rat will be explored.

Experiments by Selye (60, 61) in 1936 showed that a wide variety of noxious agents and stimuli—e.g., poisons, ionizing radiation, extreme cold, excessive exercise, surgical injury, and emotional excitement—produce a nonspecific, systemic response in the vertebrate organism. This adaptive response involves release of adrenocorticotrophic hormone (ACTH) from the anterior pituitary, and consequent release of adrenal cortical hormones, such as cortisone and 17-hydroxycorticosterone, into the blood stream. The general effect of these adrenocortical hormones is to maintain blood pressure and blood-sugar levels, thus counteracting shock effects in injury and trauma. These hormones stimulate metabolism at the cellular level[2] and, in general, their end-effects are protein catabolic, involving breakdown of protein into glucose.

There appears to be general agreement in 1958 that almost any type of disturbing stimulus induces a pituitary-adrenal response to stress in the rat, and that this reaction is essential to preserving the animal's metabolic and circulatory integrity under stress, as Engel stated in 1953 (14).

The work of Selye (60) at Montreal has indicated that prolonged stress, like cold or excessive exercise, will lead to degenerative renal and cardiovascular changes in the rat and, if sufficiently prolonged, to the subsequent death of the animal. A number of studies in the

---

[1] This material is based, in part, on a report to the National Research Council of Canada, Jan. 3, 1953, and on a paper subsequently presented to the Canadian Psychological Association, May 28, 1953, at Kingston, Ontario. Related research at the University of Toronto was made possible by Grant A.P. 20 from the National Research Council of Canada to the writer, 1952–1954, and by the friendly and continued support of E. A. Bott, Head, Department of Psychology, University of Toronto. Preparation of this paper was made possible by a grant from the Research in Humanities Fund of the University of Toronto.

[2] A stress not encountered in nature is oxygen at high pressure. Under this stress, presence of the adrenals is detrimental to survival of the rat (17). These and similar experiments (3, 16) suggest that the normal action of the adrenal cortex is to enhance metabolic processes, particularly oxidation, at the cellular level.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE

United States have, however, suggested that there are marked effects of early individual attention, including handling and petting, on growth and survival of the albino rat.

Hammett (23) in 1922 found that rats stroked or petted were much more likely to survive an operation for removal of the parathyroid gland than those that had not been handled. Bernstein (4) reported in 1952 that handled rats gained more weight than unhandled controls, and also learned mazes faster and with fewer errors.

These and other considerations suggested the hypothesis that early-handled rats would be better able to withstand severe stress than nonhandled controls. This hypothesis was tested in a series of experiments at the University of Toronto under the writer's direction.

Handling and viability. This research has shown that early handling of the albino rat increases viability under stress as an adult (7). Weininger, in two independent experiments (68, 69), found that early-handled rats, subjected to 48 hours of immobilization and food and water deprivation as adults, sustained less heart and gastrointestinal damage than unhandled controls. Long (7) found that early-handled rats survived complete food and water deprivation as adults 53.8 hours longer than matched controls, thus confirming the results of an earlier pilot study (67) by Weininger. In general, early-handled animals showed less emotional reactivity to a strange environment, and their adrenals did not become hyperreactive under stress (69). These results suggest that early handling raises the threshold for the pituitary-adrenal response to stress.

In the Toronto studies, early handling involved holding the rat loosely in the left hand against the experimenter's chest and stroking the rat "with the grain," head to tail, for 10 minutes a day, a slight modification of the method originally developed by Bernstein (5). For convenience through the remainder of this paper, this general method will be referred to as "Type I" handling. In the Toronto studies, handling was carried on for 21 days after weaning (age 21 days).

These findings have been in part confirmed by subsequent experimental work in the United States. Levine and Otis (36) at Ohio State found that rats handled prior to weaning showed significantly less mortality following food and water deprivation than nonhandled rats or rats handled after weaning. Levine (32) has further shown that early-handled mice have a higher threshold for emotional reactivity than nonhandled controls. Levine has also found recently (34) that adrenal weights of rats handled prior to weaning were significantly lower, following the stress of glucose injection, than those of controls. In the latter study, he found no significant difference in mortality under this stress between early-handled and control rats.

In these studies of Levine and his colleagues, however, early handling involved removing the infant rat from its nest, placing it in a different environment (as on a scale or in a box) and then returning it to its nest.[3] In the Levine and Otis study, for example, this picking-up-and-returning procedure probably involved no more than a minute all told, and probably no more than 12 seconds' contact with the experimenter.[4] Yet this abbreviated procedure was found to have the same protective effects under stress[5] as the Toronto method, also used by Levine and Otis in their study (36). For the sake of convenience, the kind of handling used by Levine will be referred to as "Type II" handling in this paper. It should be noted that in the Levine studies "early handling" refers in general to handling before weaning.

---

[3] In a study at Toronto, Newton (7) found that prolonged (postweaning) experience with the mother rat *increased* the infant rat's emotional reactivity as an adult, compared to handled rats, and that this extra mothering lowered the basal metabolic rate and weight for the rat, compared to controls that had not been either extra-mothered or handled.

[4] Personal communication with S. Levine, 1957.

In a replication of Weininger's experiments (68, 69), but using group-housed rather than individually-caged rats as in the Toronto studies, Stern and Taylor[6] have recently found significantly reduced gastrointestinal lesions in early-handled (Type I) rats following 48 hours of immobilization, compared to controls, but no differences in adrenal weight, heart damage, or emotional reactivity under these conditions.

An experiment by Scott (59), contrary to the findings of Weininger and Levine, showed no differences in weight, emotional reactivity, or survival under severe physical stress (massive doses of thiourea) between early-handled (Type I) and control rats. In this experiment, however, a severe stress in the form of a heat wave intervened between the early handling and the experimental tests. The effects of such intervention of stress on the growth gains usually induced by early handling are discussed below.

A number of recent studies have suggested that sensory input from other modalities may influence the effects of handling on viability. An experiment by King[7] has shown that even visual contact with others of its species has survival value for the guinea pig in infancy, when the animal is already under dietary stress (vitamin deprivation). An experiment by Shelly, Pasamanick, and Levine (62) showed that rats held loosely in a black box a few minutes a day for nearly three weeks evidenced only transient growth gains over controls during the next week, irrespective of their age at time of handling. Rats handled prior to weaning in this experiment also showed a decrease in timidity in a strange situation that disappeared after one trial. Macdonald and Teghtsoonian (38) found that three minutes of intense light stimulation produced a transient but significant weight gain in rats that had previously received maximum visual stimulation in infancy, compared to controls that also experienced the intense light as adults but that had been subjected to less visual stimulation in infancy.

These results suggest, as Thompson has hypothesized (66), that the total quantity of sensory interaction with the environment, rather than the specific nature of the sensory input, is the essential factor involved in effects of early experience on viability. If this hypothesis were correct, then the effects of early handling on viability would depend on the amount of visual stimulation concomitantly experienced by the young animal. It might be expected, in this case, that rats handled in the dark would be less resistant to stress as adults than those handled in daylight or under artificial light.

BASIC HYPOTHESES

Stress threshold. The key hypothesis of the position presented here is that early handling raises the threshold for response to stress, thus conserving the metabolic resources or adaptation energy (60) of the organism.[8] The response to stress may be differentiated into two reactions. Stimulation of the adrenal medulla by the sympathetic division of the autonomic nervous system results in release of epinephrine and nor-epinephrine into the blood stream. And, as noted above, stimulating of the adrenal cortex by ACTH from the anterior pituitary results in release of adrenocortical hormones, such as cortisone and 17-hydroxycorticosterone, into the blood stream. Both these reactions are, in general, mediated by the hypothalamus (25) and they may be considered, as Cleghorn (9) suggests, to be the two arms of a single response to noxious stimuli. It can further be suggested, with Sayers (55), that both arms of the response to stress usually occur together.

---

[5] In his experiment at Toronto, mentioned above, Long (7) found that rats that were merely picked up and held against the experimenter's chest for 30 seconds, and then returned to their cages, survived longer under total food and water deprivation than controls, but this difference was not statistically significant.

[6] Unpublished study by J. A. Stern and R. Taylor, School of Medicine, Washington University, 1957.

[7] Personal communication with J. A. King, 1957.

While the work of Selye has emphasized the pituitary-adrenal response in general in accounting for stress damage, recent investigations by Goldstein (21) and others, to be considered below, have pointed to the crucial role of the sympathetic system in tissue damage.

The basic hypothesis of this paper may therefore be restated as follows: *early handling of the albino rat reduces (a) the pituitary-adrenal and (b) the sympathetico-adrenal medulla responses to later stress.*

The evidence for each of the two sections of this hypothesis will now be considered in turn, and further relevant experiments suggested in each case.

A reduced *pituitary-adrenal output* in early-handled animals under stress is suggested by failure of their adrenals to show hyperactivity under these conditions. In two experiments, Weininger (69) found significantly lower adrenal weights in early-handled (Type I) rats than in unhandled ones, following 48 hours of immobilization and food and water deprivation. Levine (34), using a different method of handling (Type II), as noted above, found significantly lower adrenal weights in rats handled prior to weaning than in unhandled ones, following the stress of glucose injection.

A more conclusive experiment would involve comparison of the adrenal-steroid (17-hydroxycorticosteroid) blood-plasma content of early-handled and of control animals under stress, after the method of Nelson and Samuels (44). Also pertinent would be comparison of resting 17-hydroxycorticosteroid levels in early-handled and in control animals, as a further check on the hypothesis of reduced pituitary-adrenal activity in the former. A third, and critical, experiment would involve measurement of

---

[8] Conversely, this suggests that an animal with a low stress threshold would, through leakage of its reserves (which can be conceived of as the organism's neuronal, hormonal and metabolic resources) across this threshold, lose its capacity to deal with the major stresses of later life. That is to say, *early stress lowers resistance to later.* Relevant here is the finding of Jones (28, 29) that a major factor in the death rate of normal human populations from such diseases as cancer is the amount of childhood disease experienced. Early stress, including disease, can be thought of as chipping away at the body's defenses and lowering resistance to later major and lethal stress (28). On the other hand, Levine (31, 35) has hypothesized that early handling constitutes a stressful situation for the infant organism, and that such early (minimal) experience with noxious stimulation results in a greater ability of the organism to adapt to psychological and physiological stress as an adult. He has recently found, for example, that rats shocked and handled in infancy, or handled alone (the handling consisting of no more than 12 sec. of contact with the experimenter per day), showed superior weight gains and less emotionality compared to control rats that had received neither handling nor shock in infancy (33). Thus, *either* early stress lowers resistance to later, as is hypothesized here, *or* early stress enhances resistance to later stress, as Levine has suggested. This point should be experimentally decidable, using such sensitive indices of resistance to stress as the total life-span of the animal.

ACTH blood content in early-handled and unhandled adrenalectomized animals under stress, using the method of Sydnor and Sayers (65). The animal of choice for these experiments might be the guinea pig, since the resting level of 17-hydroxycorticosteroid blood plasma content in the rat is so low as to be virtually undetectable (12).

Even if it were further confirmed, however, through the experiments suggested above, that early handling reduces pituitary-adrenal output, this would by no means demonstrate that such reduction was due to a dampening of pituitary-adrenal activity. It could equally well be due to a greater capacity of the pituitary-adrenal axis in early-handled animals for the normal production of adrenal cortical steroids, as Levine has suggested (35), so that for this reason there would be less detectable output under stress.

Levine (35) has recently shown that handling (Type II) after birth induces significant increases in ascorbic acid depletion (an index of stress response) in rat adrenals under stress, and on the basis of this and other work concludes that handling accelerates maturation of the pituitary-adrenal system in the rat. In this connection, Woods (72) has shown that the wild Norway rat, with markedly greater adrenal capacity than the laboratory rat, shows no detectable adrenal ascorbic-acid depletion as an adult under such stress as low temperature, fighting, or loud noise. In this case, then, the greater capacity of the wild rat results in a reduced pituitary-adrenal output under stress, compared to the laboratory animal.

If the hypothesis that early-handled rats had, as a result of handling, greater adrenal capacity than unhandled controls, were correct, several experimentally testable hypotheses should follow.

First, this greater capacity should be evidenced in greater weight of the adrenals under normal conditions, since the greater capacity of the wild rat is reflected in adrenals proportionately two to four times the size of those of laboratory rats, the difference being in size of the adrenal cortex (50). No differences between adrenal weights of early-handled (Type I) and unhandled rats have, however, been detected prior to stress (69).

Second, this presumed greater capacity and greater normal output should be reflected in a higher adrenal-steroid (17-hydroxycorticosteroid) resting level in handled, compared to unhandled animals. But this would be precisely the reverse of the expectation if the position advanced in this paper were true, for if anterior pituitary and adrenocortical activity were permanently dampened under early handling, then the resting 17-hydroxycorticosteroid level should be *lower* than that in unhandled controls.

The question thus resolves itself into whether the obtained reduction in pituitary-adrenal output under stress of early-handled animals, as found by Weininger (69) and Levine (34), is due to a concomitant reduction in stimulation of the anterior pituitary and hence in ACTH output, as suggested in this paper, or is due to accelerated maturation and consequently greater capacity of the pituitary-adrenal system in early-handled animals, as suggested by the work of Levine (35).

A reduction of *sympathetic-adrenal medulla response* in early-handled animals under stress is suggested by the reduced emotional reactivity that has usually, but not always, been found in these animals as adults.

Weininger (69) found that early-handled (Type I) rats, introduced to a circular open field enclosed by three-foot metal walls with a 200-watt bulb hung over the center, ventured further towards the center of this arrangement than controls (two experiments), and showed more activity than controls (two of three experiments, the result being reversed in the third). Levine (32) found that early-handled (Type II) mice showed less latency in leaving their cages for food and in fighting other mice than controls. Bernstein (5) found that rats handled (Type I) prior to and during the course of the experiment showed more vigorous exploratory behavior in a maze and in their home cages than did controls, learning mazes faster and with fewer errors. Reduced emotional reactivity of handled rats could account for both their superior maze performance and greater exploratory activity.

In a recent experiment, Rosen (51) found early-handled (Type I) rats dominant in paired food competition with unhandled controls, and suggests reduced emotional reactivity of handled rats as the relevant factor. No correlation was found between weight and dominance.

On the other hand, Gertz (18) found in a recent study that early handling (Type I) of the albino rat had no effect on adult emotional behavior, as measured by frequency of defecation, number of boluses excreted, and activity in an open field.

The basic hypothesis of reduced sympathetic-adrenal medulla response of early-handled animals under stress could be tested by comparing cardiac rate and peripheral sympathetic reaction (blood pressure in the tail, for example) of early-handled and control rats under stress. It could further be tested by comparing blood and urine output of epinephrine from the adrenal medulla of early-handled and control rats subjected to noxious stimuli.

Although a number of interesting clean-up studies remain to be done in this area, it is the conviction of the writer that the point of diminishing returns has been reached in current psychological studies of the effects of the early experience of handling on the response to stress, and that the time is now ripe to enquire into the underlying neurophysiological mechanisms operating to produce the observed results.

HYPOTHALAMIC REACTIVITY

If the effects of early experience of handling on viability are to be accounted for in terms of the threshold for stress, then the effects of such experience on the neurophysiological determinants of the stress response must be shown. The writer (6) has suggested that early handling dampens posterior hypothalamic reactivity to stress. Porter (45, 46) has shown that this is the area (mammillary body and median eminence) involved in the pituitary-adrenal response to stress. Such dampening of the posterior hypothalamus would inhibit the sympathetic-adrenal medulla stress response as well as the pituitary-adrenal cortex reaction, since both of these "arms" of the response to stress are mediated by the posterior hypothalamus (25, 42, 46).

Sympathetic response. Both of these arms usually occur together, as Sayers (55) has pointed out. Virtually all conditions that release ACTH and are considered stress situations also increase sympathetic nervous activity and the secretion of epinephrine.[9]

In accounting for stress damage to the cardiovascular and renal systems, the work of Selye (60) has put the onus on the pituitary-adrenal response and the consequent production of adrenal cortical steroids. Recent investigations by Goldstein (21) and others have, however, shifted emphasis to the sympathetic division of the autonomic system as the primary agent in tissue damage under stress. For example, adrenalectomized dogs maintain a precarious balance under severe physical stress to which they would otherwise succumb, providing they have been sympathectomized or have been treated with drugs that block action of the sympathetic system (21).[9] This work of Goldstein and others suggests that adrenal cortical hormones have the function of buffering the tissues against damage from epinephrine and nor-epinephrine.

---

[9] Personal communication with M. S. Goldstein, 1956.

Without such preparation of minute blood vessels by adrenal steroid hormones, autonomic neurohumoral activity appears to result in irreversible damage, and hence probably in the terminal conditions of local haemorrhage and decreased blood pressure (21). On the other hand, interruption of the sympathetic system protects the adrenalectomized animal from stress.[9] Goldstein notes in this connection: "It is as if the price paid for a highly reactive system such as the sympathetic is to expose the tissue to damage with the steroids functioning to support the tissue in meeting the metabolic stimulation of the adrenalins."[9]

Other experiments have shown the protective effects of sympathetic blocking agents in tumbling shock (52) and in tourniquet and burn shock (40). It may therefore be suggested here that the protective effects of early handling result from just such a dampening of sympathetic activity under stress. This hypothesis could be directly tested by comparison of the relative protective effects of early handling and of sympathetic blocking agents for rats under severe physical stress, such as tumbling shock. The hypothesis could be indirectly tested by comparing the startle and peripheral sympathetic response (as in tail blood pressure) to sudden stimuli of early-handled rats and controls.

Hypothalamic balance. The major mechanism here suggested to account for the effects of early handling on viability remains, however, dampening of posterior hypothalamic reactivity (6). Inhibition of the sympathetic autonomic response, and of the pituitary-adrenal response, would be subsidiary effects of this change in hypothalamic function. This hypothesis could be tested directly by comparing electrical activity in the posterior hypothalamus of early-handled and control rats under physical, physiological, and emotional stress. For example, the electrical response to epinephrine injection could be compared in handled and control animals.

The neurophysiological mechanisms involved in such postulated posterior hypothalamic dampening have not been investigated.

Recent experimental work by Redgate and Gellhorn (15, 49) suggests, however, that balance of excitability between the anterior and posterior hypothalamus determines pulse-rate and blood-pressure levels. Coagulation or drug-induced reduction of posterior hypothalamic excitability resulted in a reduction of blood pressure and pulse rate. Coagulation of the anterior hypothalamus resulted in opposite changes.

Other research by Shimazu et al. (63) in Japan has shown that stimulation of parasympathetic hypothalamic centers in the rabbit inhibits secretion in the posterior hypothalamus of the neurosecretory granules[10] involved in the response to stress. Their findings indicate that the neurosecretory material is released into the hypothalamic and hypophysial blood vessels by stimulation of the posterior sympathetic centers of the hypothalamus, or by such stresses as wounds, burns, or poison.

---

[10] The neurosecretory cells, which are equipped to transform neural impulses into humoral activity, thus provide the mechanism for linking the two great integrating systems of the body (56).

The work of Scharrer (56, 57) and Harris (25) has indicated that it is the transmission of this material from the posterior hypothalamus to the anterior pituitary through the portal blood vessels that initiates the release of ACTH from the pituitary and thus triggers the pituitary-adrenal response to stress, particularly when the noxious stimulus is mediated by the central nervous system (as is the case for emotional stress). The important finding of the Japanese investigators, for the present purpose, is that stimulation of the parasympathetic centers suppresses the release of this neurosecretory material and causes its retention in the axon.

Those environmental stimuli that activate anterior and parasympathetic hypothalamic centers would, therefore, at the same time inhibit the posterior hypothalamic nuclei that mediate the two arms of the response to stress. The balance between activity of the anterior and posterior hypothalamus can thus be termed the threshold for stress. The more dominant the posterior hypothalamic centers over the anterior, the greater the probability that a neutral stimulus would result in a stress response, and thus the lower the threshold for this response.

To account for the effects of early handling on viability of the albino rat, it can now be suggested that early handling stimulates the anterior hypothalamus and parasympathetic centers and thus inhibits the posterior hypothalamic response to stress. If this sensory input from handling occurs early enough in the development of the organism, then the balance is permanently shifted in favor of the anterior region.

The relative dominance of one or the other of these two response systems, determined genetically or by early experience, can thus be considered the threshold for stress. That is to say, the probability that the rat will respond to a neutral environmental stimulus with a protein catabolic reaction is, in part, determined by the rat's early sample of what life is like for rats.

The central proposition of this paper is, therefore, that early handling alters the balance of hypothalamic excitability in favor of the anterior (and parasympathetic) regions, thus having the joint effects of (*a*) reducing pituitary-adrenal and sympathetic-adrenal medulla responses under stress by the mechanism noted above, and (*b*) stimulating anabolic processes, such as an increase in secretion of growth hormone from the anterior pituitary (handling has been found to increase weight and skeletal length of rats in a number of experiments, noted below), through a mechanism not yet determined.

The anticatabolic effects of early handling in reducing pituitary-adrenal and sympathetic-adrenal medulla responses under stress have been considered above. The anabolic effects of early handling will now be considered in some detail.

Anabolic effects. In respect to anabolic stimulation from early handling, the finding of Doolittle and Meade (13) that free amino-acid blood-serum content of early- and late-handled (Type I) rats shows an acute reduction, compared to unhandled controls, suggests that protein synthesis is increased, and/or protein breakdown reduced, in handled animals. This indirectly supports the hypothesis that early handling stimulates growth hormone secretion, as suggested by Harris (24).

In this connection, Mogenson et al. (41) have found that the weight gain usually induced by early handling (Type I) is nullified either by injection of cortisone (an adrenal cortical hormone) or by the stress of a high-frequency sound, both effects presumably operating to counteract growth hormone or reduce its output. Ruegamer and Silverman (54) have shown that rats handled (Type I) two weeks preceding electric shock, and with resultant gains in growth and food utilization over controls, lost these advantages when both groups were subjected to electric shock. Both groups were inferior to nonshocked controls.

This effect of stress on anabolic effects of early handling might account for the failure to obtain weight gains from handling (Type I) in the experiments of Newton (7), Long (7),[11] and Scott (59). The heat wave that intervened in the latter experiment may have served to wipe out any differences between experimental and control groups. The Newton and Long results may have been related to the exposure of their rats to cold in transit from upper New York State to Toronto in the winter, but this is only speculation. In any case, a majority of studies in the literature[12] have shown that early handling induces weight gain in the rat, and Ruegamer et al. (53) have shown this gain (from Type I handling) to be the result of better food assimilation rather than of greater consumption. This again suggests the influence of growth hormone (somatotrophin or STH).

How increased growth-hormone output could lead to enhanced survival under such specific and severe types of stress as complete food and water deprivation, as found for early-handled rats in the studies of Long (7) and Levine and Otis (36), is suggested by the relation of STH to cardiac glycogen deposition.

Studies of Adrouny and Russell have shown, respectively, that STH increases cardiac glycogen under normal conditions and that it has an important role in maintenance of cardiac glycogen stores during the emergency of fasting (1, 2). The greater survival of early-handled (Type I) adult rats (age 52 days) under food and water deprivation than of controls *of the same weight*, as found by Long (7),[11] is otherwise rather difficult to explain; but it could be accounted for on the assumption that these animals were able to maintain cardiac glycogen (a form of carbohydrate) stores longer than controls. This hypothesis could be experimentally checked by comparison of cardiac glycogen deposits in early-handled and control rats (*a*) under food and water deprivation and (*b*) under normal conditions.

If exploration of the hypothesis that early handling alters the balance of hypothalamic activity in favor of the anterior region should prove fruitful, the next logical step would be investigation of the neurophysiological determinants of such altered neuroendocrine response.

This would require, in the first instance, recording of anterior and posterior hypothalamic activity from chronically implanted electrodes, to test the specific hypothesis of anterior dominance. The specific hypothesis to be tested here would be that, in early-handled animals as compared to controls,

---

[11] Personal communication with H. G. Long, 1955.

[12] These include an original experiment by Bernstein (4), three subsequent experiments by Weininger (69, 70), three by Ruegamer et al. (53), and one by Brooker (8), all involving Type I handling. Levine found weight gains from handling (Type II) in his glucose-injection experiment (34). Both Weininger (69) and Ruegamer (53) found that early handling (Type I) increased skeletal length. In another experiment in this area, McClelland (37) found that stroking the rat (in a restraining box) with an artist's brush for the same period had exactly the same effect in inducing weight gain as personal (Type I) handling by the experimenter. This latter finding supports Weininger's view (67, 70) that the essential variable involved in handling is tactile stimulation, rather than supporting Bernstein's view (5), suggested to account for his findings on handling (Type I) and maze learning, that the essential element in handling is relationship of the animal to the experimenter.

(*a*) posterior hypothalamic excitability would be reduced, and (*b*) anterior hypothalamic excitability would be increased, under both normal and stress conditions.

The hypothesis of anterior dominance could be indirectly tested by comparing pulse rate and blood pressure of early-handled and control rats, under both normal and stress conditions. The work of Redgate and Gellhorn (49), discussed above, suggests that if the anterior hypothalamus is dominant in early-handled rats, as hypothesized, these animals should evidence, compared to controls, (*a*) a slower return of blood pressure from drug-induced hypotension, and (*b*) an increase in reflex slowing of the pulse from nor-epinephrine. The most conclusive experiment in respect to anterior dominance would, however (as suggested above), involve application of a stress agent to the organism (epinephrine, for example) and consequent recording of hypothalamic electrical activity, comparing early-handled and control rats in this respect.

MODULATION OF HYPOTHALAMUS

Amygdaloid complex. If a change in the balance of hypothalamic activity is suggested to account for effects of early handling on viability, then the question arises as to just how such a change could be effected. Recent studies by Gloor (19, 20) and Pribram (47) of the amygdaloid complex[13] and its diffuse projection system to subcortical regions suggest that one of its functions is to modulate hypothalamic excitability. Other research by Schreiner and Kling (58) has shown that removal of the amygdaloid complex reduces behavior of both savage and domestic animals to tameness and docility. The primitive function of the olfactory cortex, of which the amygdaloid complex is an element, can be thought of, after Herrick (26), as qualitative discrimination between desirable and noxious stimuli and consequent nonspecific arousal of the appropriate subcortical centers, particularly hypothalamic.

If the function of the amygdaloid complex includes discrimination between desirable and noxious stimuli, then its removal should interfere with such discrimination. Relevant here is the finding of Weiskrantz (71) that the effect of amygdalectomy on monkeys is to make it difficult for the animals to identify motivationally relevant stimuli (in the words of J. Olds), whether desirable or noxious.

A study by Lesse (30) has shown that stimulus situations involving behavioral responses of agitation, escape, attack, or defense induce in the cat a distinct electrographic pattern (high amplitude, 40–45 sec. frequency) in the lateral nuclei of the amygdaloid complex.

Since these behavioral responses are precisely those that Hess (27) has obtained from direct electrical stimulation of hypothalamic centers (also in the cat), a connection between amygdaloid and hypothalamic response to stress would seem indicated here. It should also be noted from the Lesse study that stimuli which resulted in waking or alerting responses only, and which have been found in other studies to produce large-amplitude, synchronized waves in the hippocampus (22), did not induce the amygdaloid EEG pattern typically elicited by noxious stimuli, such as electric shock.

Indirect evidence for amygdaloid

---

[13] The amygdaloid nucleus is a complex gray mass, covered by a layer of rudimentary cortex, in the dorsomedial region of the temporal lobe, near the tip of and on the floor of the inferior horn of the lateral ventricle. While it is considered to belong to the olfactory secondary centers, it has a good many nonolfactory connections, particularly somesthetic (39), and may be thought of as an olfactory thalamus or integrating center. Cf. Strong and Elwyn (64), Ranson and Clark (48).

modulation of hypothalamic reactivity comes from de Florida and Delgado's finding (10) that stimulation of the amygdaloid nucleus induces affectionate and playful behavior in a group of male cats, with effects lasting for several days. An experiment of Morgane and Kosman (43), however, contrary to the trend of the above studies, suggests that the amygdaloid complex has no effect on hypothalamic reactivity. They found that bilateral amygdalectomy in cats produced no observable changes in affective behavior, including reaction to noxious stimuli or petting.

The question remains as to how the sensory input from early handling could affect amygdaloid activity. Relevant in this connection is the study of Machne and Segundo (39), who found stimulation of the sciatic nerve and touching of any portion of the skin most effective in arousal of electrical activity in the amygdaloid nucleus. From this study, olfacto-somatic correlation appears to be a major function of this subcortical center.

The above work, taken together, suggests that the balance of hypothalamic activity, and therefore the threshold for stress, may be determined by amygdaloid mediation. The testing of this hypothesis would require removal of the amygdaloid complex by suction in experimental and matched control animals before handling, to determine whether the effects of handling on viability would thereupon be eliminated. Further, electrical recordings of amygdaloid activity during stress and under normal conditions could be made to determine any differences between early-handled and control animals.

Hippocampus. The role of the hippocampus (see Fig. 1) in modulating hypothalamic activity also requires exploration. Relevant in this respect are (*a*) the finding of Gloor (19) that hippocampal responses to electrical stimulation of the amygdaloid nucleus reach the largest amplitudes of any structure fired by the amygdala; and (*b*) a reported finding of J. W. Mason[14] that electrical stimulation of the hippocampus itself acts to suppress the pituitary-adrenal response to stress.

These findings would suggest, then, that the amygdaloid complex may have at least two pathways for hypothalamic modulation—a direct projection to the anterior hypothalamus via the stria terminalis, with direct fiber connections (19), and an indirect, multisynaptic pathway to the hippocampus, whose main efferent projection (the fornix) is to the posterior hypothalamus, in the region of the mammillary body (see Fig. 1). This is, of course, precisely the region found by Porter (45, 46) to mediate the pituitary-adrenal response to stress.

Delgado (11) found defensive reactions in monkeys following electrical stimulation of the hippocampus and fornix, and, on the basis of his own and other work, suggests the hippocampus as locus for nociceptive sensory integrations. These and other investigations impute to the hippocampus a leading role in the mediation of noxious stimuli and hence in the neuroendocrine response to stress, although no simple or clear-cut picture of hippocampal function emerges from this research.

SUMMARY

The central proposition of this paper is, therefore, that effects of early handling on viability of the albino rat are mediated by a permanent alteration in the balance of hypothalamic activity, in favor of the anterior region. This alteration results in increased growth hormone output under normal conditions, and decreased activity of the pituitary-

---

[14] Personal communication with J. V. Brady, 1957.

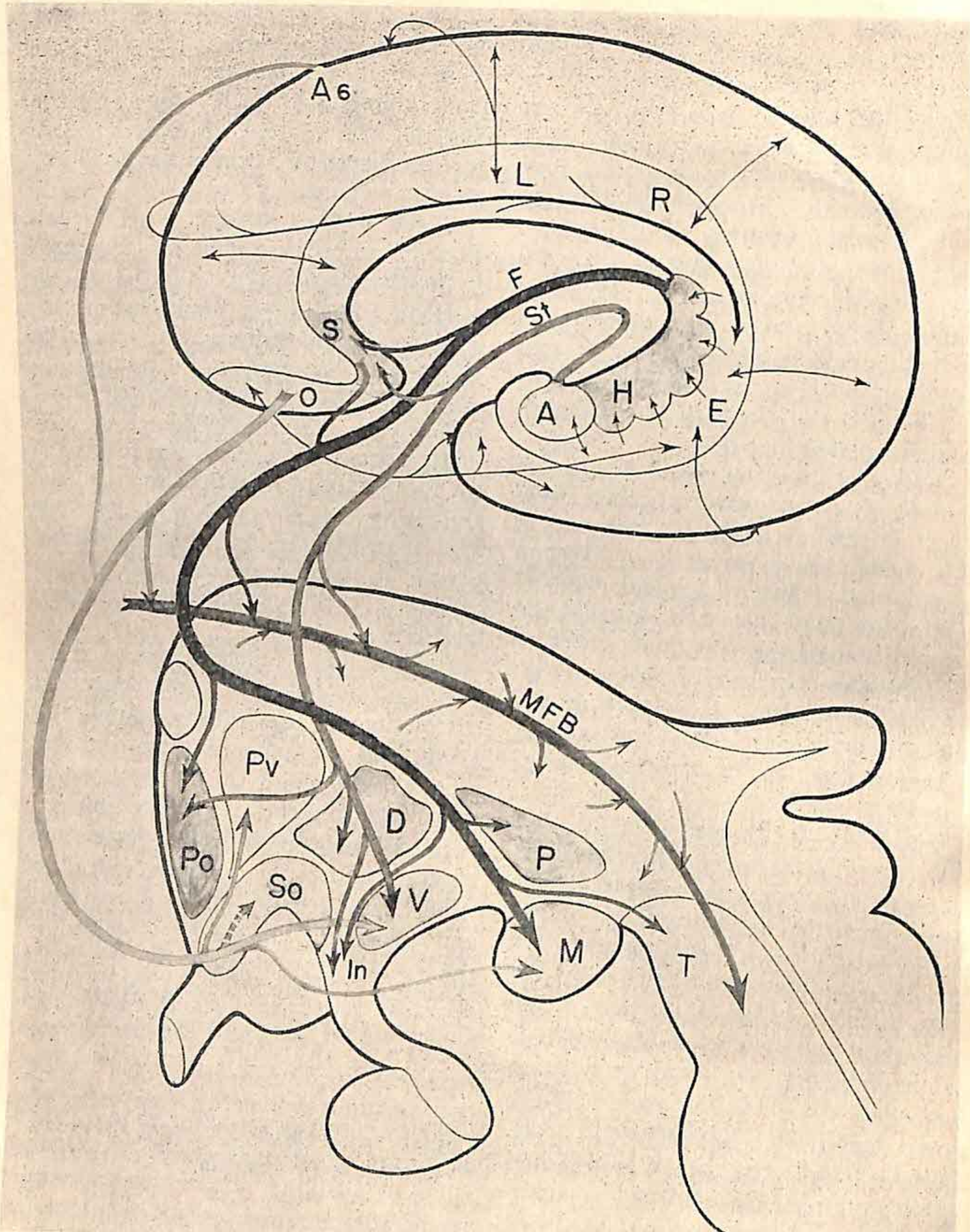


FIG. 1. Diagrammatic representation of the main telencephalo-hypothalamic connections. The arrows indicate the direction of nervous impulses in this system. The arrows in the diagram of the hemisphere outline the main associational connections at the cortical level, which are relevant for this system. (From Gloor, P. Telencephalic influences upon the hypothalamus. In W. S. Fields, R. Guillemin, & C. A. Carton (Eds.), *Hypothalamic-hypophysial interrelationships*. Springfield, Ill.: Charles C Thomas, 1956. Pp. 74–113. Courtesy of the author and Charles C Thomas, Publisher.)

A: amygdala. A6: Area 6 of frontal lobe.
D: dorsomedial nucleus of hypothalamus.
E: entorhinal area (pyriform lobe).
F: fornix.

adrenal cortex and sympathetico-adrenal medulla systems under both normal and stress conditions. In turn, this alteration in hypothalamic balance is itself the result of a change in amygdaloid complex activity, arising from the sensory input from early handling.

Experimental tests of derivative hypotheses have been suggested in the course of this presentation.

REFERENCES

1. Adrouny, G. A., & Russell, Jane A. Growth hormone and myocardial glycogen. *Fed. Proc.*, 1954, **13**, 1. (Abstract)
2. Adrouny, G. A., & Russell, Jane A. Effects of growth hormone and nutritional status on cardiac glycogen in the rat. *Endocrinology*, 1956, **59**, 241–251.
3. Bean, J. W. Reserpine, chlorpromazine and the hypothalamus in reactions to oxygen at high pressure. *Amer. J. Physiol.*, 1956, **187**, 389–394.
4. Bernstein, L. A note on Christie's "Experimental naiveté and experiential naiveté." *Psychol. Bull.*, 1952, **49**, 38–40.
5. Bernstein, L. The effects of variations in handling upon learning and retention. *J. comp. physiol. Psychol.*, 1957, **50**, 162–167.
6. Bovard, E. W. A theory to account for the effects of early handling on viability of the albino rat. *Science*, 1954, **120**, 187.
7. Bovard, E. W., & Newton, D. G. Systematic early handling and prolonged experience with the mother as developmental variables in the male albino rat. In N. R. Brewer (Ed.), *Proceedings of animal care panel, 1955*. Chicago: Animal Care Panel, 1956. Pp. 67–74.
8. Brooker, H. The effect of the amount of early handling on weight gain in the albino rat. Unpublished master's thesis, Univer. of Toronto, 1955.
9. Cleghorn, R. A. Discussion of F. L. Engel, General concepts of adrenocortical function in relation to the response to stress. *Psychosom. Med.*, 1953, **15**, 571–573.
10. de Florida, F. A., & Delgado, J. M. R. Lasting effects on behavior evoked by cerebral stimulation of the cat. *Amer. J. Physiol.*, 1955, **183**, 592.
11. Delgado, J. M. R. Cerebral structures involved in transmission and elaboration of noxious stimulation. *J. Neurophysiol.*, 1955, **18**, 261–275.
12. Done, A. K., Ely, R. S., Raile, R. B., & Kelley, V. C. Species differences in circulating 17-hydroxycorticosteroid concentrations. *Proc. Soc. exp. Biol., N. Y.*, 1952, **81**, 667–669.
13. Doolittle, R. F., & Meade, R. The effect of gentling on some psychological and physiological phenomena. Paper read at Eastern Psychological Association, New York, April 13, 1957.
14. Engel, F. L. General concepts of adrenocortical function in relation to the response to stress. *Psychosom. Med.*, 1953, **15**, 565–573.
15. Gellhorn, E. *Autonomic imbalance and the hypothalamus.* Minneapolis: Univer. Minnesota Press, 1957.
16. Gerschman, Rebecca, Gilbert, D. L., Nye, S. W., Nadig, P. W., & Fenn, W. O. Role of adrenalectomy and adrenal-cortical hormones in oxygen poisoning. *Amer. J. Physiol.*, 1954, **178**, 346–350.
17. Gerschman, Rebecca, & Nadig, P. W. Stress and oxygen poisoning. *Fed. Proc.*, 1953, **12**, 50–51. (Abstract)
18. Gertz, B. The effect of handling at various age levels on emotional behavior of adult rats. *J. comp. physiol. Psychol.*, 1957, **50**, 613–616.
19. Gloor, P. Electrophysiological studies on the connections of the amygdaloid nucleus in the cat. Part I: The neuronal organization of the amygdaloid projection system. *EEG clin. Neurophysiol.*, 1955, **7**, 223–242.
20. Gloor, P. Electrophysiological studies on the connections of the amygdaloid nucleus in the cat. Part II: The electrophysiological properties of the amyg-

---

H: hippocampus.
In: infundibulum.
L: limbic (cingulate) gyrus.
M: mammillary body.
MFB: medial forebrain bundle.
O: orbitofrontal cortex.
P: posterior hypothalamus.
Po: preoptic region.
Pv: paraventricular nucleus.
R: retrosplenial and posterior cingulate region.
S: septal area (subcallosal region).
So: supraoptic nucleus.
St: stria terminalis.
T: tegmentum.
V: ventromedial nucleus.

daloid projection system. *EEG clin. Neurophysiol.*, 1955, **7**, 243–264.

21. Goldstein, M. S., Ramey, E. R., Fritz, I., & Levine, R. Reversal of effects of stress in adrenalectomized animals by autonomic blocking agents. *Amer. J. Physiol.*, 1952, **171**, 92–99.
22. Green, J. D., & Arduini, A. A. Hippocampal electrical activity in arousal. *J. Neurophysiol.*, 1954, **17**, 533–557.
23. Hammett, F. S. Studies of the thyroid apparatus. *Endocrinology*, 1922, **4**, 221–229.
24. Harris, G. W. *Neural control of the pituitary gland.* London: Arnold, 1955.
25. Harris, G. W. Hypothalamic control of the anterior lobe of the hypophysis. In W. S. Fields, R. Guillemin, & C. A. Carton (Eds.), *Hypothalamic-hypophysial interrelationships.* Springfield, Ill.: Charles C Thomas, 1956. Pp. 31–42.
26. Herrick, C. J. The function of the olfactory parts of the cerebral cortex. *Proc. nat. Acad. Sci., Wash.*, 1933, **19**, 7–14.
27. Hess, W. R. *Diencephalon. Autonomic and extrapyramidal functions.* New York: Grune & Stratton, 1954.
28. Jones, H. B. Demographic consideration of the cancer problem. *Trans. N. Y. Acad. Sci.*, 1956, Series II, **18**, No. 4, 298–333.
29. Jones, H. B. A special consideration of aging process, disease and life expectancy. In J. H. Lawrence and C. A. Tobias (Eds.), *Advances in biological and medical physics, 4.* New York: Academic Press, 1956. Pp. 281–337.
30. Lesse, H. Electrographic recordings of amygdaloid activity during a conditioned response. *Fed. Proc.*, 1957, **16**, 79. (Abstract)
31. Levine, S. A further study of infantile handling and adult avoidance learning. *J. Pers.*, 1956, **25**, 70–80.
32. Levine, S. Emotionality and aggressive behavior in the mouse as a function of infantile experience. *Amer. Psychologist*, 1957, **12**, 410. (Abstract)
33. Levine, S. Infantile experience and consummatory behavior in adulthood. *J. comp. physiol. Psychol.*, 1957, **50**, 609–612.
34. Levine, S. Infantile experience and resistance to physiological stress. *Science*, 1957, **126**, 405.
35. Levine, S., Alpert, M., & Lewis, G. W. Infantile experience and the maturation of the pituitary adrenal axis. *Science*, 1957, **126**, 1347.
36. Levine, S., & Otis, L. The effects of handling before and after weaning on the resistance of albino rats to later deprivation. *Canad. J. Psychol.*, 1958, **12**, 103–108.
37. McClelland, W. J. Differential handling and weight gain in the rat. *Canad. J. Psychol.*, 1956, **10**, 19–22.
38. Macdonald, G. E., & Teghtsoonian, R. Weight gain in the albino rat as a function of early visual experience. Paper read at Canadian Psychological Association, Toronto, June 7, 1957.
39. Machne, Xenia, & Segundo, J. P. Unitary responses to afferent volleys in amygdaloid complex. *J. Neurophysiol.*, 1956, **19**, 232–240.
40. Millican, R. C., & Stohlman, E. F. Relative effectiveness of certain drugs against shock produced in mice from tourniquet and burn trauma. *Amer. J. Physiol.*, 1956, **185**, 195–200.
41. Mogenson, G. J., McMurray, G. A., & Jaques, L. B. Effects of stress and administration of cortisone on weight gain in gentled rats. *Canad. J. Psychol.*, 1957, **11**, 123–127.
42. Morgan, C. T., & Stellar, E. *Physiological psychology.* (2nd ed.) New York: McGraw-Hill, 1950.
43. Morgane, P. J., & Kosman, A. J. Alterations in feline behavior following bilateral amygdalectomy. *Fed. Proc.*, 1957, **16**, 90. (Abstract)
44. Nelson, D. H., & Samuels, L. T. A method for the determination of 17-hydroxycorticosteroids in blood: 17-hydroxycorticosterone in the peripheral circulation. *J. clin. Endocrinol.*, 1952, **12**, 519–526.
45. Porter, R. W. Alterations in electrical activity of the hypothalamus induced by stress stimuli. *Amer. J. Physiol.*, 1952, **169**, 629–637.
46. Porter, R. W. Hypothalamic involvement in the pituitary-adrenocortical response to stress stimuli. *Amer. J. Physiol.*, 1953, **172**, 515–519.
47. Pribram, K. Concerning the neurophysiological correlates of limbic system stimulation. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Houston, Texas: Univer. Houston Press, in press.
48. Ranson, S. W., & Clark, S. L. *The anatomy of the nervous system.* (9th ed.) Philadelphia: Saunders, 1953.
49. Redgate, E. S., & Gellhorn, E. Relation of anterior hypothalamic excitability to

cardiovascular reflexes. *Amer. J. Physiol.*, 1955, **183**, 654. (Abstract)

50. RICHTER, C. P. Domestication of the Norway rat and its implications for the problems of stress. In H. G. Wolff, S. G. Wolf, Jr., & C. C. Hare (Eds.), *Life stress and bodily disease.* Baltimore: Williams & Wilkins, 1950. Pp. 3–18.
51. ROSEN, J. Dominance behavior as a function of early gentling experience in the albino rat. Unpublished master's thesis, Univer. of Toronto, 1957.
52. ROSS, C. A., & HERCZEG, S. A. Protective effects of ganglionic blocking agents on traumatic shock in the rat. *Proc. Soc. exp. Biol., N. Y.*, 1956, **91**, 196–199.
53. RUEGAMER, W. R., BERNSTEIN, L., & BENJAMIN, J. D. Growth, food utilization, and thyroid activity in the albino rat as a function of extra handling. *Science*, 1954, **120**, 184–185.
54. RUEGAMER, W. R., & SILVERMAN, F. R. Influence of gentling on physiology of the rat. *Proc. Soc. exp. Biol., N. Y.*, 1956, **92**, 170–172.
55. SAYERS, G. The adrenal cortex and homeostasis. *Physiol. Rev.*, 1950, **30**, 241–320.
56. SCHARRER, E. The general significance of the neurosecretory cell. *Scientia*, 1952, **46**, 177–183.
57. SCHARRER, E. Neurosecretion and the anterior pituitary in the dog. *Experientia*, 1954, **10**, 264–266.
58. SCHREINER, L., & KLING, A. Rhinencephalon and behavior. *Amer. J. Physiol.*, 1956, **184**, 486–490.
59. SCOTT, J. H. Some effects at maturing of gentling, ignoring or shocking rats during infancy. *J. abnorm. soc. Psychol.*, 1955, **51**, 412–414.
60. SELYE, H. *Stress.* Montreal: Acta, 1950.
61. SELYE, H., & FORTIER, C. Adaptive reactions to stress. In H. G. Wolff, S. G. Wolf, Jr., & C. C. Hare (Eds.), *Life stress and bodily disease.* Baltimore: Williams & Wilkins, 1950. Pp. 3–18.
62. SHELLY, CAROLYN, PASAMANICK, B., & LEVINE, S. The effects of infantile experience upon adult behavior in the albino rat. Paper read at American Psychiatric Association, Chicago, May 16, 1957.
63. SHIMAZU, K., OKADA, M., BAN, T., & KUROTSU, T. Influence of stimulation of the hypothalamic nuclei upon the neurosecretory system in the hypothalamus and the neurohypophysis of rabbit. *Med. J. Osaka Univer.*, 1954, **5**, 701–727.
64. STRONG, O. S., & ELWYN, A. *Human neuroanatomy.* Baltimore: Williams & Wilkins, 1948.
65. SYDNOR, KATHERINE L., & SAYERS, G. A technic for determination of adrenocorticotrophin in blood. *Proc. Soc. exp. Biol., N. Y.*, 1952, **79**, 432–436.
66. THOMPSON, W. R. Paper read in panel discussion, Effects of early experience. Canadian Psychological Association, Toronto, June 7, 1957.
67. WEININGER, O. Mortality of albino rats under stress as a function of early handling. *Canad. J. Psychol.*, 1953, **7**, 111–114.
68. WEININGER, O. Physiological damage under emotional stress as a function of early experience. *Science*, 1954, **119**, 285–286.
69. WEININGER, O. The effects of early experience on behavior and growth characteristics. *J. comp. physiol. Psychol.*, 1956, **49**, 1–9.
70. WEININGER, O., MCCLELLAND, W. J., & ARIMA, R. K. Gentling and weight gain in the albino rat. *Canad. J. Psychol.*, 1954, **8**, 147–151.
71. WEISKRANTZ, L. Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *J. comp. physiol. Psychol.*, 1956, **49**, 381–391.
72. WOODS, J. W. Differences in adrenal response to adverse conditions in wild and domesticated Norway rats. *Fed. Proc.*, 1953, **12**, 159. (Abstract)

(Received February 4, 1958)

Psychological Review
Vol. 65, No. 5, 1958

FUNDAMENTAL DIMENSIONAL PROPERTIES OF THE OPERANT[1]

THOMAS F. GILBERT

University of Georgia

At the outset of any inquiry into behavior, and prior to the introduction of experimental variation, the investigator prepares to perform two classes of operations. In the first place he locates and circumscribes, in space and time, the subject matter with which he is concerned. Secondly, he designates some specific dimension or dimensions along which quantities of his subject matter can be established. It is the purpose of this paper to describe the simplest set of operations that can be used to establish quantities of behavioral subject matter. However, the acceptability and meaning of any description of a fundamental dimensional set will depend, in part, upon what unit of subject matter is chosen for purposes of quantification. To put it simply, *how* one sets about to measure largely depends upon *what* one sets about to measure.

Attempts at systematic designation of behavior dimensions have been hampered by the confusion about just what events enter into the subject matter which the dimensions are to describe. If we look at what measures behaviorists actually use we see an indefinite number of dimensions. Latency, resistance to extinction, habit strength, amplitude, speed, accuracy, threshold, frequency, reaction-potential, aggression, mels, VTE, rate, and anxiety are only the beginning of the list of the most common "dimensions." If we reflect upon these many different measurements we should find that many of them are, for the sake of measurement, reducible to others. For an example, resistance to extinction, as a "measure" of reflex strength, is established by measuring frequency, amplitude, or latency after reinforcement has been withdrawn. Resistance to extinction is a phrase we use to describe changes in behavior, measured along some dimension, as a function of a certain kind of experimental variation. In short, resistance to extinction designates a functional relationship. Threshold is another case in point. We do not "measure" threshold, but we measure the frequency with which one says, "yes, I see the light," and we describe how the frequency of this behavior is functionally related to quantitative changes in the experimental variable. We then, sometimes, worry our heads over the fact that this functional relationship is nonlinear and does not interest the two zeroes simultaneously. Seldom, in science, are functional relations between two variables linear, and seldom does the relationship intersect the variables at the point of origin. This is no ground for leading us to conclude that basic psychological measurement dimensions do not have the characteristics of ratio scales.

It is the first business of science to describe functional relations. It carries out this business by introducing experimental variation and measuring the concurrent changes in the subject matter. In order to describe these functional relations, some operations of measuring the subject matter necessarily are established *prior to* the introduction of the experimental variations. In any science, functional relationships exist in indefinite numbers, while measurement dimensions do not.

---

[1] This research was supported, in part, by the Graduate School, University of Georgia.

The operations of measuring behavior and the operations of establishing functional relations must be independent of each other, lest confusion is to arise. The procedures we use to describe behavior are established at the beginning of inquiry and cannot change unless the inquiry itself is redirected. Descriptive procedures are required to be independent of experimental variation, since these procedures are utilized for describing the effects of the experimental conditions. This is not to say that the *quantities* obtained from descriptive procedures must be independent of the experimental variables; rather, the *procedures* will be independent if the dependence of the quantities is to have meaning. A set of measurement dimensions is established for the purpose of describing functional relationships and will, therefore, not include the very functional relationships to be described.

A second difficulty in dimensional analysis lies in the confusion about what is meant by dimensional independence, or unidimensionality. The precise distinction that it is necessary to make at the outset is that each member of the fundamental set represent an independent procedure. This does not mean that the outcomes of measurement along different dimensions will be uncorrelated. It simply means that there is nothing in the procedures themselves that necessitates correlation. When we ask how many dimensions a house possesses, we do not first obtain data on the house; rather we first decide along what dimensions houses are measurable, regardless of the particular quantities obtained in the measurement of any one house. We say that a house is multidimensional simply because there is more than one procedure by which we can measure it. A dimension implies an operation for obtaining data independent of the particular quantities obtained on one or more occasions. However, in the study of behavior, dimensional analysis is frequently thought to be a procedure for ordering data already obtained independently of the methods used to obtain it. Thus, we let the unidimensional characteristics of a set of data hinge upon the data's being uncorrelated with some other set of data. It is not surprising that "unidimensionality" has been hard to come by. It is a well known fact that the physical dimensions of mass and extension are frequently correlated in nature. Nevertheless, we think of them as independent dimensions because they are operations of measurement that do not necessitate correlation. They are, as I shall say, *logically* independent, although not necessarily *empirically* independent.

The analysis that follows does not, then, refer to operations for establishing quantitative functional relationships, nor of operations for establishing the degree of empirical correlation between a set of measures already obtained. It refers, instead, to the simplest set of operations for obtaining measurements of the behavior of organisms. It seems clear that those who investigate behavior typically investigate that class of behaviors that Skinner has called operant (8, p. 65). Skinner's well known definition is suitable for the purposes of this paper. The dimensions set forth here will, I think, be relevant to the class of respondents as well, although I am not fully satisfied that this is so.

Once the behavior is specified as an operant, it still has no empirically determined quantity. A definition of the operant denotes operations that yield its qualitative characteristics and establish it as subject matter. There remains the problem of describing the simplest possible operations that can be performed in the determination of empirical quantity. These operations, once they are specified, I shall call primary dimensions. Primary dimensions will be those operations of measuring that yield quantities that are of an irreducible simplicity. These quantities will also possess the characteristics of ratio scales.

Now these quantities will all be derived by the physical operations of the measurement of time and space, and can be reduced to these operations; but in being so reduced, their behavioral convenience will be lost. They are irreducible only in the sense that further reduction takes the quantities out of the realms of behavioral discourse. So long as we regard the phenomena in the realms of psychological investigation as physical and extensible, as I shall do here, measurements in time and space are called for. As Bridgman (1) has said, dimensions are matters of convenience so long as they are compatible with the facts and possess the characteristics of ratio scales. Time measurements are themselves extensible, and can be measured by the distance a hand travels over a clock face. This does not remove the fact of the convenience of time as a primary dimension in physics.

The dimensional operations I shall specify here are not given as the absolute and "real" character of behavioral data, but are offered as convenient methods for assigning quantities to behavioral phenomena. The final test of their fundamental properties will be in the consequences that accrue from their use in the laboratory. As I shall describe them, they will be designations of certain operations that can be carried out by the investigator, who has established an operant as his subject matter. They will not refer to specific characteristics of behavioral data, but to means by which these characteristics can be described. It should also be obvious that, if these dimensions can be designated as more fundamental than other procedures of measurement, these other procedures must be reducible to the dimensions described here.

THE DIMENSIONS

The task of outlining the nature of the dimensions that I consider primary may be facilitated by imagining a simple laboratory situation. A human subject is brought into a room and seated before a panel on which is mounted a plunger that can be pushed against a spring resistance. The plunger is so built that, if it is released at all, it will return to its normal or starting position before it can be pushed again. The plunger is mounted on a swivel so that the direction of the push is variable. A cumulative recorder is attached to the plunger so that any movement is recorded. The subject is instructed to push the plunger as often as he likes, and told he may earn money for his work. He is told to begin when the experimenter turns on a light and that he can stop working whenever he likes. The fixed reference space, in this case, is the plunger, and the operant is plunger pushing.

Temporal Dimensions

Latency. The first elementary measurement operation is that of recording the time between the opportunity for an operant and its initial occurrence. The measure is stated in time units, and can be easily obtained from a cumulative recorder or a timer that goes into operation when the plunger is made available and cuts off when the plunger is first pushed.

Tempo. Once an operant occurs it is frequently repeated, and there is a unitary quality about the sequence of operants as they are being repeated. For example, the child who pedals a bicycle evidences no break between pedaling through one revolution and another, and one operant leads smoothly into another. The rate of this continuously ongoing operant is tempo. In the plunger-pushing example, tempo can be easily measured as the slope of the cumulative curve during those times in which *S* is working continuously. This measurement is made quite easily where these slopes are linear, which is more often the case than not. A decision must be made about what is to be considered continuous or discontinuous; however, this decision is taken care of in the consideration of the next dimension.

The measure of tempo as the rate at which the operant is being emitted, exclusive of any breaks in its emission, is dependent upon the duration of the individual operants in the run. The tempo of a single operant is its duration. In order to distinguish between the duration of the operant and the duration of the time in which the organism is emitting operants, I shall borrow a term from physics. We may refer to the duration of a single instance of an operant as the *period* of that operant.

Perseveration. The third temporal dimension can be seen where we regard the amount of time the organism spends, after latency, at the prescribed task. If we regard the total time from the occurrence of the first operant to the completion of the last, the *proportion* of this time in which operants are actually occurring is perseveration. Examples of this type of variation are numerous. For example, after some continuous plunger pushing or bicycle pedaling, a person will pause and engage himself in some other task before continuing with the original one. Just how long the pause must be before we decide that the task has been discontinued is an arbitrary matter. This question creates very little difficulty in normal laboratory situations, however. In the plunger example, perseveration is measued by a timer that is in continuous operation while *S* is pushing the plunger. If *S* pauses for as much as five seconds, a delay relay shuts off the timer until *S* resumes work. Perseveration can also be calculated directly from cumulative tracings. The five-second pause as a criterion for deciding that *S* has quit the task is arbitrary, but is derived, in this case, from observations of how long it takes the slowest *S* to complete the cycle of one plunger push.

Duration. The fourth temporal dimension is the total time from the initiation of the first operant to the completion of the last, independent of the amount of work done within this interval. The designation of the completion of a sequence of operants will vary from investigation to investigation. It may be in the form of completing a task such as filling up a ditch, reaching some criterion of satiation, or returning to operant-level of output. In the plunger example, it is simply the time from the end of latency until *S* pushes a switch, thus indicating his desire to quit.

Figure 1 is a schematic presentation

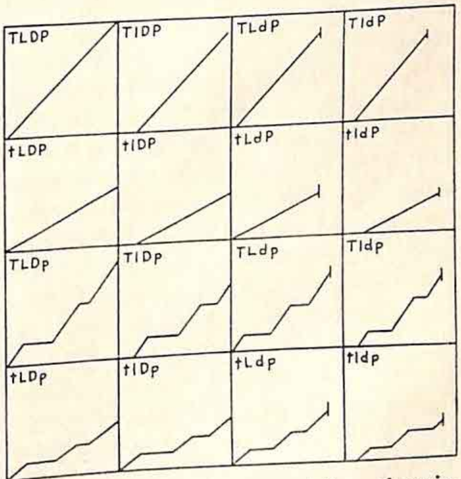


Fig. 1. A schematic presentation of variations along four temporal dimensions, illustrating the independence of each. The figures are cumulative recordings. Upper case letters represent: fast tempo (T), short latency (L), long duration (D), and high perseveration (P). Lower case letters represent: slow tempo (t), long latency (l), short duration (d), and low perseveration (p).

of hypothetical results from the plunger-pushing experiment. The figure represents several cumulative tracings, the ordinate being a frequency scale and the abscissa a time scale. As can be seen readily, the total work output may vary as a function of each of these dimensions, and the variation in each dimension is logically independent of variations in any other dimension.

Output

Total output data will be a function of each of these four temporal dimensions. The dimensions combine in the following manner to yield a total output quantity (where it can be assumed that tempo is linear):

$$R = KTPD - L,$$

where R is output in frequency; P is perseveration in proportion of total time less latency; D is duration in seconds; L is latency in seconds; T is the slope of the cumulative recording where the animal is at work; and K is a correction for different units of measurement of frequency on the ordinate and seconds on the abscissa. KT is the number of operants occurring in a second during the times in which the animal is actually working. It should be noted that the spatial dimensions also enter into the determination of total output. Here we have considered the case in which these dimensions are held at some fixed value.

In the special case in which the task is of some fixed length, duration naturally will be determined by perseveration and tempo. In this event, time rather than output measures are commonly used. It can be seen that time is a complex measure, determined by the operation of several dimensions.

Spatial Dimensions

A most obvious and common measure of an operant in space is what has been called amplitude or magnitude. However, the behavioral concept of amplitude may involve two independently measurable characteristics. For this reason I have chosen two names that, while not unusual, are seldom used in behavioral description.

Intension. The operation that is usually indicated by the word "amplitude" is the measurement of the amount of movement that takes place in some specified part of the organism's topography. This measure of operant amplitude I shall call intension, and it is taken when we regard the amount of movement that some specified part of the animal makes with reference to some other space. In the plunger experiment, the measure is simply made by recording the distance through which the plunger moves from the normal position. Intension must not be confused with the next measure of amplitude.

Extension. Independently of the amount of movement of a specified part of the animal, we may measure the amount of the animal involved in the movement. A simple illustration of the distinction between intension and extension is made by imagining a hand laid palm down on a table surface. We may regard the distance through which a finger (or fingers) may be lifted (intension), or we may regard, independently of the distance through which they are moved, whether one or more fingers are lifted (extension). Russell et al. (5) have made a somewhat similar distinction in an analysis of quantitative procedures. I believe that what I call extension is similar to what they have called "quantal intensity." A precise terminological comparison is hardly feasible, since our methods of analysis differ rather sharply. The measurement of extension is difficult to make in many behavioral situations, including the plunger apparatus. We may determine whether the plunger is pushed with one hand or two, or we may make arbitrary divisions of the topography and measure the involvement of these divisions by muscle action-potentials.

Direction. The third spatial dimension is so well known as to require little elaboration. This dimension is established by the measurement of the direction of movement with respect to some fixed reference space. The operation by which this measure is taken from the swivel-mounted plunger should be obvious. Accuracy, error, and "preference" are typical expressions of the direction dimension.

TWO EXPERIMENTAL EXAMPLES OF THE ANALYSIS

Such an analysis as the one offered here should have empirical consequences to establish its worth to an empirical science. I will briefly describe a few representative results from two investigations that may serve to argue for the more widespread usage of this analysis as well as to indicate the empirical properties of several of these dimensions.[2]

The first experiments represented procedures similar to those of Guttman (3). Although no attempt was made to replicate Guttman's work, our data are not very different from his findings. Essentially we allowed fully trained rats to receive .05 ml. of a sucrose solution from a tube upon pressing a bar (Guttman used .005 ml.). One of the major experimental variables was the concentration of sucrose in solution (4, 8, 16, and 32%). An important feature of the apparatus was the fact that the animal received a drop directly into his mouth as he pressed the bar, and he never had to move away from the bar. In many Skinner boxes, the animal moves away from the bar and the operants cannot occur with the continuity required for the measurement of tempo.

From rather fast-moving cumulative recorders, the temporal dimensions listed above were measured. Figure 2 summarizes and schematizes the findings. Each of the temporal dimensions is plotted as a separate function of the sucrose concentrations. Figure 2E represents the total output as a function of sucrose concentration, and the general form of this curve is similar to that found by Guttman. However, we can see that the total output is analyzable into four contributing dimensions, each being a different function of the experimental variable.

Tempo. This is an invariant, remaining the same throughout the range of concentrations (Fig. 2C). It was also discovered that tempo did not change as the animal approached satiation; a fact that is supported by earlier findings in this laboratory, as well as by the data of Skinner (7), Young (9), and Sidman and Stebbins (6). This invariant character of tempo seems to be true for both PR and continuous schedules. The generalization is that, when the animal works at a task, he works at a tempo that is characteristic of him and that is unaffected by the nature of his reinforcers or the extent of his deprivation.

Another characteristic of tempo seems to be its moment-to-moment invariance. In our laboratories we investigated the consistency of tempo from day to day and from time to time within a day's run. Reliability coefficients, represented by linear *r*s, are in the neighborhood of .98.[3] There is great variability in tempo between animals, and negligible variability within animals from time to time.

---

[2] The data of the first of these experiments were obtained in collaboration with Lloyd Hitchcock, Jr. (4).

[3] This work was done with W. T. James of the University of Georgia, and reported to the Southern Society of Philosophy and Psychology, Atlanta, April, 1954. James is primarily responsible for the design of the liquid reinforcing apparatus which was used.

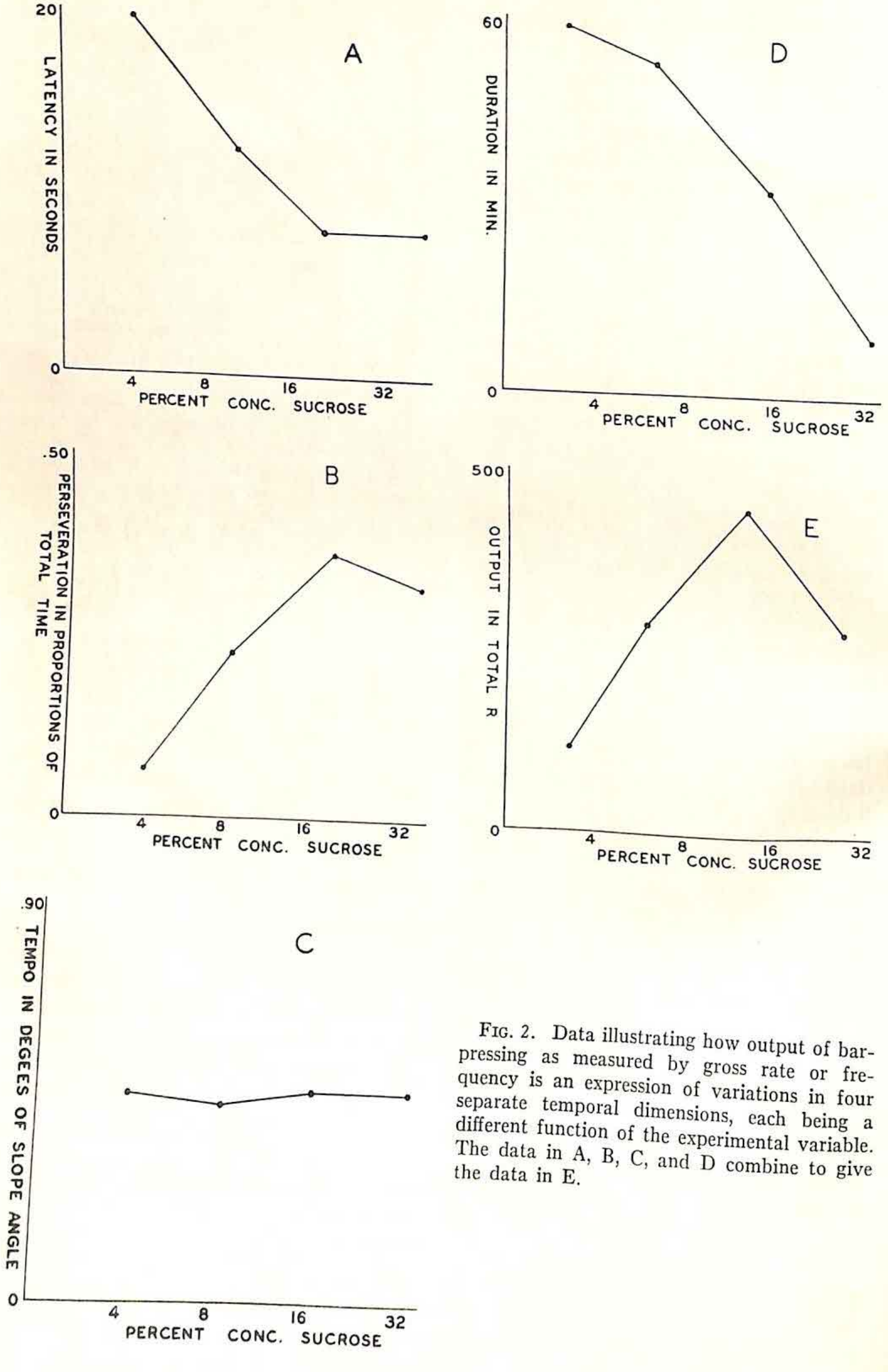


FIG. 2. Data illustrating how output of bar-pressing as measured by gross rate or frequency is an expression of variations in four separate temporal dimensions, each being a different function of the experimental variable. The data in A, B, C, and D combine to give the data in E.

Latency. In the sucrose investigation, it was discovered that latency is a negatively accelerated, decreasing, monotonic function of reinforcing concentrations (Fig. 2A). This is a finding supported by too many investigations to need mentioning. The well known variability in latency measures was also observed here.

Perseveration. This was found to be an increasing, negatively accelerating, monotonic function of sucrose concentration (Fig. 2B). Actually, variations in perseveration determine the greater part of the variations in total output, and seem to be the most sensitive and reliable indices of performance variability. The reliability of these measures is quite high. As the animal approaches satiation, he perseveres less and less at the task, and this fact probably determines the negative acceleration of most work curves.

Duration. In the sucrose reinforcement experiment the animals were allowed to work a fixed time, so that all animals did not reach a reasonable criterion of satiation. However, estimates from these data combined with other data taken in our laboratory suggest the following conclusions: Duration is a positively accelerating, decreasing, monotonic function of sucrose concentration. The greater the concentration of sucrose, the earlier does the animal become satiated (Fig. 2D). However, as deprivation is increased, duration becomes greater—a fact that should be disturbing to anyone who would point to the similarity of the effects of drive and reinforcement on the form of total output curves.

A brief summary of data from the human plunger-pushing apparatus, described above, may serve to indicate some of the generality of these procedures and their outcome. The human subject is requested to push the plunger to work for points. He receives one point after the first 200 plunger pushes, and then no more. He is allowed to work as long as he wishes. Figure 3A represents the cumulative recordings of this work for three different *S*s. Figure 3B represents the curves of a typical rat working for different amounts of sucrose. As can be seen, tempo is invariant in all *S*s, and tempo, perseveration, and duration all serve to determine the final output.[4]

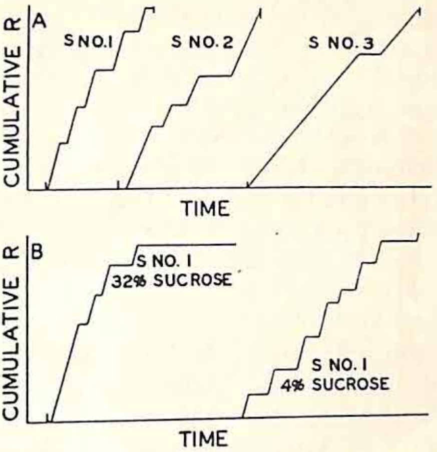


FIG. 3. Sample cumulative tracings illustrating the constancy of tempo and the way in which total output is determined by variation in different dimensions. A represents three human *S*s pushing a plunger to earn grade points. B represents one rat working for two different concentrations of sucrose.

SOME IMPLICATIONS OF THIS ANALYSIS

Consideration of Statistics

It is commonly believed that statistical analysis is a necessity in the behavior laboratory. A few investigators, such as Skinner, have presented persuasive arguments to the contrary. In spite of these arguments, however, data are often so variable as to appear understandable only by way of statistical analysis. It is usually assumed that this great variability is produced by uncontrolled and unknown "random" conditions in the laboratory, and that statistical analysis allows us to estimate and account for these "random" effects. I shall not deny the grounds for the validity of this assumption; however, I believe that this is a much overworn argument. I contend that the compound nature of much of our measurement is an important source of data variability. In the experiment described above, we obtained curves shaped like those of Guttman (3). However, with twelve animals a statistical analysis yielded no significant *F*s. Guttman had to use several dozen animals to obtain *F*s that were only moderately significant. When we analyzed the perseveration scores separately, there was no question about statistical significance. Such significance was found using the analysis of variance, but it was quite obvious without the use of statistics. This was more remarkably the case with tempo scores.

---

[4] Skinner (7) once demonstrated the use of a water operant. The curves he presents are similar to those here, and the constancy of tempo can be readily seen, although his technique was slightly different. He rejected the water operant partly on the grounds of the irregularity of the curves obtained. The irregularity indicated breaks in the work periods.

If, as I have contended, an animal's tempo is invariant under a wide range of experimental conditions, it can only serve to attenuate differences in total output scores produced by experimental variation. If an animal spends most of his time at a task, and if he works at a tempo unaffected by the experimental conditions, the tempo will be the greatest determinant of the total output measure. The output score will then consist of a component reflecting not experimental variations but great interindividual differences. In the language of statistics this could only serve to increase the "error term."

Misleading Characteristics of Total Output Measures

To exemplify the greater need for the type of analysis I have proposed, an experiment by Fink and Patton (2) will be scrutinized. These investigators studied the effects of changes in the population of stimuli on the decrement of conditioned drinking behavior. Rats were taught for five days to drink from a burette. On the sixth day various kinds and degrees of stimulus changes were made in the drinking box, and measurement was made of the water intake during the first two minutes following the stimulus changes. It was concluded that "the greater the number of stimulus components changed, the greater the response decrement" (2, p. 27). The findings of this experiment are interpreted by the authors as giving support to the S-R contiguity theory. The authors note that their measure could reflect either rate or amplitude, and point that that the Estes-Burke behavior models have not yet made differential formulations for rate and amplitude. However, it is important to notice that rate is not, itself, a simple measure. Should we assume that their data reflect no changes in amplitude, the interpretation would remain quite difficult. The decrement in the drinking reflex could be due to changes in latency. Perhaps the animals were "startled." If the decrement did not reflect latency changes, but decreasing tempo, the meaning of the findings should be considerably changed. Similarly, changes in perseveration or duration would be reflected in the total output during the two minutes and would deserve different interpretations. If "amplitude" changes account for the decrement, it must be shown whether these changes were in intension or extension. Precision in mathematical and verbal formulations of theory must always follow precision in the measurements on which they are based. Otherwise, the precision is specious indeed.

The Fink and Patton experiment is only illustrative. The amount of possibly misleading output data appearing in behavioral journals would require a separate review. Experimentation on work may serve as another general example. Data from work experiments are notoriously variable and the uncontradicted conclusions about this field are few. Output measures are favored, and they reflect each of the proposed dimensions as they combine to yield a total score. Precise answers about, say, optimum rest periods will await the study of the effects of these rest periods at different tempos, perseverations, durations, intensions, and extensions. Similarly, questions about optimum tempo will have answers dependent upon the relationship of tempo to perseveration, duration, extension, etc.

A Word About "Mental Tests"

Intelligence test scores are compounds, not only of measurements of different operants, but of various dimensions of several of the separate operants. Thus, the Wechsler-Bellevue Digit Symbol test score will vary as the examinee perseveres during the task, as a function of his tempo, as a function of latency, as a function of direction (accuracy) and perhaps as a function of intension and extension when he uses the pencil. Only duration is fixed. Two identical Digit Symbol scores may have entirely different meanings as a result. Whatever questions concerning the construction of ratio-scales of measurement exist, one thing seems clear: If scores are constructed from the undifferentiated combinations of seven different ratio scales, no one may reasonably hope that the result will be a ratio scale, regardless of what further operations are performed upon these scores. No two IQs are ever likely to mean the same things when such measurements are uncritically combined. Even when the tests seem to meet the criteria of ratio scales, as the Stanford version of the Binet-Simon so nearly does,[5] the comparisons are still meaningless in any but the crudest sense. This is not because the tasks required are different, since they can all be considered as belonging to the same task class; it is because the tasks are measured differently or by compound procedures. Equal things, measured differently, are not equal.

SUMMARY

Seven dimensions of an operant were described as fundamental in the sense that other operations of behavior measurement are reducible to these seven and the seven are not further reducible within the realms of behavioral discourse. Certain operating assumptions were outlined as a basis for the analysis.

The seven dimensions are (*a*) *latency*, defined as the time between opportunity for an operant and its initial occurrence; (*b*) *tempo*, defined as the rate of emission of a continuously ongoing operant, or the period of a single operant; (*c*) *perseveration*, defined as the proportion of the time from the occurrence of the first operant to the completion of the last in which operants are actually occurring; (*d*) *duration*, the total time from the initiation of the first operant to the completion of the last; (*e*) *intension*, the amount of movement that takes place in an operant involving some specified part of the organism's topography; (*f*) *extension*, the degree of the topography involved in an operant; and (*g*) *direction*, i.e., of the organism's movement with respect to some fixed reference point.

Empirical characteristics of several of these dimensions were described. An experiment, partly modeled after Guttman's (3) work on sucrose as a reinforcer, was performed to indicate how each of several dimensions may be a different function of an experimental variable, and combine to give a peculiar output function. Implications of the analysis were discussed, with particular reference to the pitfalls of uncritically combining the dimensions.

---

[5] E. E. Cureton, personal communication. Cureton has shown how this test has nearly equal SDs from the ages of 3 to 13, and how the scale, if extrapolated, places zero M.A. at birth.

REFERENCES

1. BRIDGMAN, P. W. *Dimensional analysis.* New Haven: Yale Univer. Press, 1931.
2. FINK, J. B., & PATTON, R. M. Decrement of a learned drinking response accompanying changes in several stimulus characteristics. *J. comp. physiol. Psychol.,* 1953, **46**, 23–27.
3. GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.,* 1953, **46**, 213–223.
4. HITCHCOCK, L. The performance of an operant in relation to the concentration of sucrose used as reinforcing agent. Unpublished M.S. thesis, Univer. of Georgia, 1956.
5. RUSSELL, W. M. S., MEAD, A. P., & HAYES, J. S. A basis for the quantitative study of the structure of behavior. *Behaviour,* 1954, **6**, 153–205.
6. SIDMAN, M., & STEBBINS, W. C. Satiation effects under fixed schedules of reinforcement. *J. comp. physiol. Psychol.,* 1954, **47**, 114–116.
7. SKINNER, B. F. Thirst as an arbitrary drive. *J. gen. Psychol.,* 1935, **12**, 40–65.
8. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.
9. YOUNG, P. T., & RICHEY, H. W. Diurnal drinking patterns in the rat. *J. comp. physiol. Psychol.,* 1952, **45**, 80–89.

(Received October 14, 1957)

Psychological Review
Vol. 65, No. 5, 1958

THE DEFINITION AND ANALYSIS OF PERCEPTUAL LEARNING[1]

JOACHIM F. WOHLWILL

University of California, Berkeley

In recent years considerable attention has been given to the question of how, and to what extent, learning may influence perception. Most frequently this question has been answered by recourse to intervening constructs such as organizations, hypotheses, assumptions, schemata, etc., as mediators of changes in the perception of a given stimulus as a function of experience. This approach has been aptly termed by Gibson and Gibson (16) the "enrichment theory," since it takes "for granted a discrepancy between the sensory input and the finished percept and . . . aim[s] to explain the difference. . . . The development of perception must then necessarily be one of supplementing or interpreting or organizing" (16, p. 33). Against this view the aforementioned authors oppose an interpretation of their own, emphasizing progressive differentiation of the stimulus as the essential element in perceptual learning. According to this "specificity theory,"

> . . . we learn to perceive in this sense: that percepts change over time by progressive elaboration of qualities, features, and dimensions of variation. . . . Perceptual learning, then, consists of responding to variables of physical stimulation not previously responded to (16, p. 34).

This thesis, as elaborated in their paper, did not long remain unchallenged. In an extended critique by Postman (30), it was argued that the specificity theory begged the question of learning, i.e., of *how* the increased differentiation of the stimuli was mediated so as to result in differential responses. In this connection Postman maintains that "the changes in response are part and parcel of the problem of perceptual learning. The need to account for changes in response inevitably endows the problem of perceptual learning with an associative component" (30, p. 442). But even taking the specificity hypothesis as a descriptive generalization of the *results* of perceptual learning, Postman argues that it is of rather limited generality: for instance, it cannot account for the perception of signs and symbols.

It seems impossible to evaluate these criticisms without some more clearly specified criterion of perceptual learning than is provided by either of the parties to this dispute. Thus the question whether perceptual learning is a matter of attaching responses to stimuli depends on the nature of the S-R relationship that is to be learned. In a typical paired-associate learning task—e.g., learning to identify a set of unfamiliar faces by name—it is evident that the association of the correct response plays a major part in the learning process. This is not to deny, of course, the possible role of perceptual processes, but it is in the very nature of the case that these two aspects cannot be disentangled. In a situation, on the other hand, in which the subject makes judgments of "same" or "different," the attaching of responses need hardly be considered, as will be shown below. As for Postman's second criticism, one is inclined to agree that in the learning of meaningful signs and symbols a stress on the differentiation of characteristics of the stimulus is generally beside the point. The question, however, is whether a theory of perceptual learning need, or in fact should, encompass this kind of task at all. The answer again hinges on the criterion for classifying a learning task as perceptual.

---

[1] This paper was written in part during the author's tenure of a NSF postdoctoral fellowship at the University of Geneva. The author is indebted to Jan Smedslund for a critical reading of the manuscript.

It is thus the intent of this paper to propose such a criterion, which will help us in evaluating the arguments brought forth in this controversy and, more important, enable us to undertake a more systematic analysis of the nature of perceptual learning in a variety of specific situations.

A CRITERION FOR PERCEPTUAL LEARNING

In formulating this criterion, our main aim will be to exclude from consideration situations which necessarily involve the formation of new stimulus-response associations, since in such situations the responses cannot be used to index changes on the perceptual side. Accordingly, the basic ingredient for our definition of perceptual learning will be the use of responses which have previously been established to a class of stimuli or to a stimulus relationship; if the stimuli which are to be discriminated represent instances of this class or relationship, the response to it *need not* then be newly associated. Perceptual learning might thus be regarded as the development of a transfer of a previously learned set of responses to a new set of stimuli, the possibility for this transfer inhering in the physical characteristics of the stimuli. In other words, each stimulus to be discriminated must be classifiable, *qua* physical stimulus, into the class of stimuli to which the corresponding differential response is preattached, and similarly for stimulus relationships.

By this definition, then, paired-associate learning and similar tasks are left outside the province of perceptual learning situations. We will return later to the problem of assessing the role that perceptual learning may play in such tasks. For the moment it may be noted that this restriction likewise excludes the learning of signs and symbols, so that Postman's second criticism, referred to above, is avoided.

In a more positive sense, the criterion proposed above has the advantage of confronting us with a clear-cut issue concerning the nature of perceptual learning. An associational basis is no longer forced upon us; instead, two alternatives now emerge. On the one hand, the transfer which the situation provides for may take place in fact. In this case perceptual learning may be interpreted in terms of progressively more effective abstraction of those features of stimulation which represent the basis for the transfer of the correct response. On the other hand, this response may be established through association, direct or mediated, with *any* differentiating aspect of the respective stimuli, through differential reinforcement.

Although the first alternative leaves open the precise nature of the mechanism by which the transfer comes to be established, it entails a variety of experimental consequences serving to distinguish it from the second alternative. One of these may be stated at the outset: if associational mechanisms are involved, learning should take place only as a function of external reinforcement —i.e., the subject's response must have some consequence informing him of its correctness or aptness. But such reinforcement need not intervene in learning which conforms to the first alternative, since this is essentially a generalization phenomenon. Other implications of this distinction will be brought out below, in the discussion of specific experimental problems.

This brings us back to the interpretation of perceptual learning offered by Gibson and Gibson (16). Their specificity theory, based on stimulus differentiation instead of mediational processes, may be classed with our first alternative (perceptual learning as transfer, through abstraction of relevant aspects of the stimulus), while the "enrichment theory" to which they oppose themselves conforms to the second (perceptual learning through association).

The above two formulations of perceptual learning may, then, be said to differ with respect to their conceptions of what is learned. As the Gibsons note in their reply to Postman (17), it is precisely this question which their specificity hypothesis is meant to clarify. In order to explore this question more specifically, however, it will be essential to examine a variety of particular situations. Indeed, we shall see presently that the answer to this what-is-learned question depends to a considerable extent on the particular discriminative judgment made by the subject.

ANALYSIS OF PERCEPTUAL LEARNING SITUATIONS

It is evident that, by the very nature of the criterion adopted, perceptual learning must be considered relative to a particular set of available responses. For the purposes of the following survey, therefore, the various kinds of situations will be differentiated according to the type of discriminative response demanded of the subject.

On this basis three broad categories of tasks may be distinguished: discrimination of identity vs. difference, quantitative discriminations, and qualitative discriminations. Each of these last two can be further subdivided into a general and a specific type. Thus we have, with respect to the quantitative category, judgments of relationship and ordered magnitude judgments; and, with respect to the qualitative category, categorization and identification.

Type I: Identity vs. Difference

This type represents undoubtedly the most basic kind of discrimination; it is applicable to any stimulus material whatever, simple or complex, familiar or unfamiliar. For this reason it lends itself ideally to the study of perceptual learning; yet the Gibsons, in their exploratory experiment illustrating their specificity hypothesis (16), seem to be the only ones to have realized the possibilities of this form of discriminative judgment for the investigation of perceptual learning.

The study of this type of differentiation requires only the prior acquisition of the differential responses to identity and difference in a group of stimuli.[2] These responses are, furthermore, completely nonspecific with respect to the aspects of the stimuli to which they refer: any differentiating characteristic among the set of stimuli constitutes a potential basis for the response. The abstraction of such a characteristic is therefore a necessary and sufficient condition for the transfer of the response, so that intervening associative processes may be considered beside the point. For this type of learning, in fact, reinforcement may be altogether unnecessary, as was the case in the experiment by the Gibsons just cited. In a sense their hypothesis is thus true here by definition: in this situation perceptual learning necessarily consists in making finer discriminations among the sources of variation present in the stimuli.

[2] For nonverbal subjects, these responses could be established through preliminary training by the matching-from-sample or oddity techniques, directed at the formation of learning sets. The success which monkeys (cf. 19) and even birds (24, 27) have had in solving problems of this type argues itself for the rudimentary nature of this form of discrimination.

It is possible, however, to interpret this hypothesis in a more specific sense, i.e., as postulating discrimination along continuous dimensions of variation, as opposed to the isolation of discrete cues. Thus, in learning to differentiate a set of faces, does the observer pick out individual details (a crooked nose, a double chin, etc.), or does he discriminate dimensional differences (width of mouth, ratio of distance between eyes to that between forehead and chin, etc.)? The importance of this question lies in the fact that the generalizability of the learning is dependent on which of these two forms of learning occurs. If the discrimination is dimensional, the learning should transfer to a new sample of faces, whereas little or no transfer is to be expected from the isolation of specific cues.

It appears likely that the complexity of the stimuli will determine in large measure the extent to which dimensional aspects of the stimuli will be discriminated. Thus the nonsense shapes generated according to information-theory principles (6) suggest themselves as promising materials for the study of this question.

Type II: Judgments of Quantitative Relationships

Here we find represented the work on the reduction of simple differential thresholds through practice, which accounts for a considerable portion of the perceptual learning literature (cf. the review by E. Gibson [10]). But this category includes in addition a variety of more complex problems, such as the perceptual constancies, illusions, etc. The essential element is the presentation of stimuli varying on some physical dimension, in accordance with one of the classical psychophysical methods (other than that of single stimuli), or some other procedure involving comparative judgment. An additional requirement, following from our definition of perceptual learning, is that the subjects have previously applied the response categories to the dimension which is manipulated. Although this stipulation rarely presents a problem with verbal subjects, it does create a judgment situation somewhat different from that in Type I.

In order to illustrate the characteristic features of this type of judgment, let us consider a subject who has been asked to judge the relative distance of two objects. The first point to be noted is that, in contrast to Type I situations, the differential responses refer now specifically to a particular physical dimension. But in the subject's prior experience with this dimension, a complex of sensory information, both dimensional (e.g., linear perspective, gradients of texture) and nondimensional (e.g., interposition), has necessarily been correlated with the physical dimension of distance. One might say, in fact, that the concept of distance, connotatively considered, is linked to this complex of stimulation.

Accordingly, in order to show improvement due to practice in his distance judgments, our adult subject need only scan the stimulus field more effectively, so as to abstract some of this correlated information. This view would seem to represent the position taken by the Gibsons on the problem of perceptual learning. Essentially descriptive though it remains, such an interpretation may yet be strengthened against the charge of circularity, to the extent that the experimenter succeeds in specifying the particular variables of stimulation which are being differentiated in this situation. The exploration of such variables is of course the aim of the "global psychophysics" advocated by Gibson (15).

Are other explanations for the learning observed in this situation possible?

If this learning took place in the absence of external reinforcement, it is difficult to conceive what they might be, except perhaps for general warm-up or learning-to-learn effects. The associational alternative, in any case, would scarcely be applicable, since the formation of S-R associations would seem to presuppose some reinforcing agent.

If, on the other hand, reinforcement is given, the associational hypothesis clearly becomes relevant. Learning may consist, then, in a direct association of the correct responses with some arbitrary differentiating cues in the stimulus objects. One readily testable implication of this formulation may be pointed out: it should be possible to establish increasingly nonveridical judgments as easily as increasingly veridical ones, simply by reversing the pattern of reinforcements. Any difference in favor of veridical learning would support the first explanation, in terms of direct transfer through improved differentiation of relevant stimulus variables.

Judgments of distance represent, of course, only one kind of task, and a relatively complex one, among the many different problems included under Type II. It is relevant to ask, therefore, whether the above discussion is applicable generally to other problems of this type. The answer depends on the complexity of the stimulus correlates of the physical dimension selected, in the context of the situation in which the judgment is to be made. Where these correlates are as univariate, as dimensionally simple as the physical dimension itself (e.g., the loudness of pure tones in a noiseless chamber, the brightness of flat surfaces under constant illumination), there appears to be no opportunity for the kind of perceptual learning postulated by the specificity hypothesis. It is perhaps significant, in this connection, that in her extensive review of the perceptual learning literature (10), E. Gibson cites no instance of improvement in differential thresholds for these intensitive dimensions. For most supposedly simple stimulus dimensions, however, the sensory input is of a somewhat higher order of complexity. Particularly is this true in the case of vision, where there are almost always some relationships within the stimulus field that vary concomitantly with the dimension being manipulated. For example, in as relatively bare a stimulus field as that represented by the lines for the horizontal-vertical illusion, variation of the vertical will be accompanied by changes in the (virtual) triangle formed by the vertical, the horizontal, and the line joining the extremities of these two. As for the still more complex, multivariate stimulus fields generally encountered outside the laboratory, it is apparent that the opportunity for the utilization of a variety of correlated information in the proximal stimulus is often considerable—a point which the Gibsons are at pains to emphasize (**16**).

Type III: Ordered Magnitude Judgments

If, for the relational responses of Type II tasks, we substitute an ordered set of responses serving to identify the magnitude of a corresponding set of points on a physical scale, we arrive at a Type III task, commonly known as scale learning. Two aspects of scale learning may be distinguished: the absolute and the relative. It will become apparent that, when only the second of these aspects is considered, the general nature of the problem resembles in many ways that of a Type II task. But since scale learning has typically been studied in regard to its absolute aspect, let us start by examining it.

Scale learning in the absolute sense consists in the identification of the points on the stimulus scale in terms of their absolute physical magnitude. The

Bureau Edni. Psy. Research
DAVID HARE TRAINING CO

difficulty with this aspect, from the standpoint of the present approach to perceptual learning, is that one can rarely be confident that the differential responses have been learned beforehand. Thus one would first have to establish them through training, as E. Gibson and her associates have done for judgments of distance over the ground (11, 12). It was demonstrated in these studies that this training transfers to subsequent judgments of a new set of points on the same field (11), and even to points on a different terrain (12). This evidence is, however, of equivocal relevance to the question of perceptual learning, since the training consisted in the establishment of the responses, and there was no opportunity for learning during the transfer series itself. Gibson and Bergman (11), as a matter of fact, recognize the ambiguity of their results, for they point out that the transfer they found may reflect chiefly the improved conceptualization of the stimulus scale, through a more accurate notion of the unit of measurement.

An alternative possibility, discussed in part by these authors, appears of greater interest for the study of perceptual learning: the subjects may have become more accurate in their discrimination of the relative ordering of the points, and of the relative distances between them. This brings us to the relative aspect of scale learning, which, although it has not been explicitly studied thus far, lends itself much more readily to an analysis in conformance with our definition of perceptual learning.

For this aspect requires only that the subject have learned a set of responses conforming to the ordinal, interval, or ratio character of the stimulus scale. He would then be in a position to apply this response set to the perceived ordering of the points on the stimulus scale, and, where applicable, to the perceived metric of the scale itself. We would thus be interested primarily in the regression, linear or otherwise, of the judgments on the objective values, and the problem of perceptual learning would reduce to the approach of this regression to unity.

The above approach rests, of course, on the assumption that the subject's numerical responses reflect the metric character of the perceived magnitudes. Garner, Hake, and Eriksen (9) have questioned the validity of this assumption, pointing out that the subjects may use the numbers, not in terms of their arithmetic relationships, but rather as identifying responses, or perhaps as a set of ordered responses. The former would be true particularly where the subject has a record of prior associative experience in the application of numbers to different magnitudes on a dimension, as in the case of size, where particular values are frequently associated with individual objects.

This is certainly a valid criticism, but it may not be critical in the study of perceptual learning. For the question remains how to account for the *improvement* in a set of judgments in a learning task. It may be, of course, that the opportunity for the association of numbers with individual points on the scale is provided by the learning situation itself. An example of this kind may be seen in an earlier study of distance estimation by Gibson and Smith (13), in which it was observed that the subjects merely learned to associate specific, isolated cues on the terrain corresponding to a given distance with the verbal response which was reinforced for that distance. But it should be possible to spot learning of this direct associative character by means of transfer tests involving other points on the scale, or a change in context, or discrimination on other, correlated dimensions. (In this particular study, the authors were interested in transfer to size-at-a-distance judgments. Not surprisingly under the circumstances, none was found.) If, in terms of the relative aspect of the scale, such transfer is found, or if learning occurs without benefit of reinforcement, it would seem to justify the inference that the subjects actually did learn to perceive more accurately the relative magnitude of the points on the scale.[3] This learning is then presumably attributable to improved discrimination of relevant stimulus variables, so that the problem of the nature of perceptual learning becomes essentially equivalent to that discussed under Type II situations.

Since the discussion so far has centered almost exclusively on the scale of distance over the ground, it is important to note that this scale is not only particularly rich in associated variables of stimulation, as noted in the previous section, but represents a rather special case, insofar as all of the points on the scale are continually exposed to the subject's view. This property clearly contributes to the subject's chances of effectively ordering a set of designated distances. For other visual scales (area, brightness) one might similarly expose the scale in its entirety, but for dimensions on which only one or a few magnitudes at a time can be exposed (pitch, weight), the situation is probably not comparable. At the very least it would be essential with such scales to present anchoring stimuli, identified as such, at frequent intervals during the learning task. Otherwise the subject's response scale might easily "drift," thereby rendering the results quite equivocal with respect to the relative aspect of the scale judgments.

Type IV: Categorization

This type of perceptual learning is concerned with the placing of unfamiliar stimuli into a set of nondimensional categories previously learned by the subject. The differential responses are thus the concepts representing the categories; they are preassociated to a class of stimuli sharing certain physical characteristics, which are present in less readily discriminable form in the stimuli to be categorized.

This problem apparently still awaits experimental attack, but one way of approaching it might be as follows: sets of nonsense shapes varying around a prototype—as described in the recent experiment by Attneave (5)—would be selected to establish an artificial set of categories; other shapes constructed so as to be further removed from their respective prototypes would then be presented for categorization. The question here is whether the subjects would, with practice, succeed in making the necessary discriminations along the variables on which the categories could be differentiated, or whether the categorization of the new stimuli would proceed by direct association, through reinforcement. This problem is clearly relevant to the study of concept formation: does a child, in learning to classify a duck as a bird, actually discriminate the features of the stimulus object which relate it to other birds with which the child is already familiar, or does he merely associate the term "bird" with this new stimulus? The evidence on this kind of question is lacking, however. To be sure, a number of concept-formation studies have manipulated variables relating to the stimulus (e.g., 7, 26), but the interest has been limited to the effect of such variables on the speed of acquisition of the concepts, rather than on the kind of discrimination of the stimuli achieved.

---

[3] It is conceivable, to be sure, that the learning may reflect the gradual abandonment by the subject of preferential response habits, e.g., the tendency to respond in round numbers. This reservation points to the importance in perceptual learning experiments of eliminating differential response strengths in advance; cf. the discussion of Type V tasks below.

Type V: Identification

The identification of stimuli by responses specific to them is one of the more frequently studied types of tasks in the investigation of perceptual learning. Yet, according to our criterion, this very specificity of the responses creates an apparent problem: if the responses are pre-established to the specific stimuli which are to be discriminated, these cannot, in themselves, be unfamiliar. But how can there be any room for learning, then? Indeed, the only way in which learning can enter into such a situation at all is through the presentation of familiar stimuli under conditions which make recognition difficult. But under these circumstances the problem of perceptual learning becomes a rather different one, involving factors which relate to the general perceptual conditions, or else to the identifying responses, rather than to the particular stimuli presented.

This point is well illustrated in some of the studies employing methods of "impoverished stimulation" to make recognition difficult (through exposure of stimuli tachistoscopically, in the periphery of the eye, under dim illumination, over a background of noise, etc.—cf. the review by E. Gibson [10]). Thus, in the study by Howes and Solomon (21) on the recognition of words in the tachistoscope, the words in the second half of the list were identified more readily than those in the first half, although the words in the two halves were not the same. This finding reflects thus a generalized practice effect, attributable to mechanisms of a peripheral nature perhaps, such as receptor adaptation or changes in focusing responses.

A factor of a rather different sort, relating to the response system, may also manifest itself in this type of situation. Where a considerable number of response alternatives are available to the subject, these may not be equal in strength initially. Learning may then consist primarily in the gradual elimination of these preferential response habits. This appears to be the explanation, for instance, for at least part of the learning which occurred in Seward's study of letter recognition (31), in which subjects tended to limit their responses to but a portion of the alphabet at the beginning.

While one may devise ways of controlling for these extraneous factors (e.g., response strength can be equated by making all required responses available to the subject throughout the task), it is doubtful whether studies employing impoverished stimulation methods will yield much information of value to the study of perceptual learning. There are, however, a variety of other ways of making recognition difficult. For instance, the usual brightness relationships in the stimulus may be reversed, by the use of photographic negatives. Alternatively, objects may be presented from unfamiliar angles or in unfamiliar contexts, and, more particularly, in motion (cf. 14). If this kind of task gives rise to a learning process, it may be of considerable interest in its own right for the study of perceptual learning. But it is apparent that learning in a situation of this sort is not easily handled either in terms of the differentiation of variables of stimulation or in terms of associative processes. This type of task has, however, been investigated much too little to justify speculation as to the nature of possible learning effects.

THE PERCEPTUAL ASPECT OF PAIRED-ASSOCIATE LEARNING

In view of the important role often played by perceptual processes in the association of new responses to a set of stimuli, a criterion for the investigation of perceptual learning which seemingly eliminates this problem from consideration may well be objected to as unduly restrictive. But it has been argued here only that it is not useful to study the paired-associate task itself as a perceptual learning situation, due to the confounding of perceptual and response variables. One may still apply the analytic framework developed in this paper to this problem, by examining the *results* of such a learning process with respect to the differentiation of stimuli that has taken place. This procedure may help, in fact, in clarifying some issues in the study of paired-associate learning.

Consider the work on the problem of stimulus predifferentiation, recently reviewed by Arnoult (4). This problem is concerned with the generally positive transfer effects obtained by establishing verbal responses to a set of stimuli, prior to the association of a different set of responses (typically motor) to the same stimuli. As Arnoult points out in his review, the nature of this phenomenon remains obscure, and there are at present not even sufficient grounds for choosing between a stimulus-differentiation interpretation and other alternative explanations which have been proposed to account for it. It seems plausible to relate this equivocal conclusion to the ambiguity inherent in the situation itself, with respect to the role of perceptual processes.

This role could be studied more directly by testing for the differentiation of the given set of stimuli on one or more relevant perceptual dimensions, before and after the association task. One would then be able to decide what has been learned about the stimuli during the course of the paired-associate learning, and thus to correlate the extent and kind of differentiation achieved with the learning in the transfer task.

The design just proposed actually resembles that used in the studies by Arnoult (3) and Hake and Eriksen (18). Parts of these two studies concerned the effect of the association of letters of the alphabet to sets of nonsense shapes on the subsequent recognition of these shapes on a same-different basis. But this particular basis is not very revealing, since the mastery of the association task already implies that the stimuli have been differentiated from one another. The fact that in both studies the results were actually negative may thus seem rather puzzling—indeed, they have been cited as negative evidence on the question of stimulus predifferentiation. The paradox disappears, however, when we realize that in the recognition task of both studies new, arbitrarily chosen stimuli were introduced along with the old ones. Thus in order to expect transfer in this situation one would have to assume that the subjects had learned more about the training stimuli than was in fact required to differentiate them.

"EXTRINSIC" VS. "INTRINSIC" CUES IN PERCEPTUAL LEARNING

The approach to the problem of perceptual learning presented in this paper is intended chiefly for the analysis of restricted experimental situations, involving perceptual judgments under controlled practice. Yet some of the major issues in perceptual learning refer to the process by which the individual, during the course of his development, comes to perceive the world as he does. In fact, the contrast between "enrichment" and "differentiation" processes delineated by Gibson and Gibson appears to be directed at conceptions of perceptual development in this more general sense.

But since the acquisition of the responses to stimuli must form an integral part of the process of development, it becomes difficult in this case to isolate the perceptual aspect of what is termed "perceptual learning." Thus, insofar as this long-term learning is concerned, the Gibsons' opposition between enrichment and differentiation may not be entirely warranted. There are undoubtedly two mutually dependent processes going on simultaneously: increasing differentiation of stimuli on the one hand, and development of intervening structures, such as proposed in "enrichment" theories to mediate the resulting discriminations, on the other hand.

Yet in one respect it seems possible to formulate a fairly sharp difference between the major empiricist interpretations of perceptual learning and the Gibsons' specificity theory. This difference, which parallels closely that proposed earlier between direct transfer and response association as bases for perceptual learning, concerns the type of features of the stimulus which the individual learns to discriminate.

One may distinguish, in effect, between those aspects of the stimulus which are extrinsic and those which are intrinsic to a particular judgment. The Gibsons are clearly concerned with the intrinsic stimulus correlates of perception—intrinsic in the sense that these aspects of the proximal stimulus invariably reflect the particular aspect of the physical environment which is being judged, by the very nature of the sensory apparatus and of the physical energies impinging on it. For example, the edges of a road seen from different heights will necessarily converge towards the horizon at differing angles (cf. 15, p. 140). Thus the angle of convergence constitutes an intrinsic correlate of perceived height. These correlates are therefore independent of experience with specific objects or in specific environments.

In contrast, the functionalists, as represented by the transactional school (22) or by Brunswik's probabilistic functionalism (8), do not differentiate between such intrinsic information and that conveyed by objects or environmental features of a more extrinsic type with which the subject may have come into contact. Thus, for Brunswik, perception is a matter of the utilization of cues of varying degrees of "ecological validity," whatever their origin. Similarly, the assumptions of the transactionalists are based in part on experience with particular meaningful objects (e.g., the typical sizes of ping-pong and billiard balls [20]) and in part on experience with universal aspects of stimulation (e.g., the successive stimulation of adjacent points on the retina due to a moving object [34]).[4]

The question which emerges from these two opposing viewpoints is the following: are extrinsic and intrinsic cues functionally equivalent, or are there differences, quantitative or qualitative, in the roles played by these two kinds of information? With respect to the basic stimulus dimensions of the physical world, at least, such differences are surely not altogether implausible. It seems conceivable, at any rate, that the nature of the perceptual mechanism of the organism is such as to predispose him automatically to respond to information of the intrinsic type, on account of its very reliability.

In order to answer this question, the two types of cues must be put into conflict with each other, as in Slack's (32) study of the size estimation of normal, over-, and undersized chairs in the presence of the ordinary cues to distance. It was found that the judgments were displaced somewhat towards the assumed size of a chair, although the true sizes of the stimuli accounted for the major portion of the variance in these judgments by far. But this study tells us little about the actual learning processes that may be reflected here, as we are only shown its end result.[5]

---

[4] In Ames' (2) rotating trapezoidal window effect, both of these kinds of cues are invoked. There is, on the one hand, experience with windows and rectangular objects, and on the other hand, experience with perspective transformations in general. It is of interest that Allport and Pettigrew (1) have recently thrown doubt on the former of these experiential factors by showing that amount of experience with windows and rectangular forms is to a large extent unrelated to the perception of this illusion.

A more direct line of attack on this problem would be to juxtapose extrinsic and intrinsic cues in a perceptual learning experiment. Thorndike, in exploratory and rather sketchily documented work (33, pp. 77 ff.), has in effect employed this method, demonstrating, for instance, that through arbitrary association with an irrelevant cue (shape), nonveridical judgments of area could be induced. The relative importance of the two types of information could similarly be assessed by correlating them in a perceptual learning task, and rendering the extraneous cue irrelevant, or even reversing it, once the discrimination had been established. Of perhaps even greater interest might be the introduction into a perceptual learning experiment of a readily discriminated extraneous cue under probabilistic reinforcement. Subsequent tests with this cue removed would indicate to what extent the subject relies on highly visible though relatively unreliable cues, in preference to possibly quite subtle but perfectly reliable sensory information. This kind of investigation might thus be expected to throw light on the relative merits not only of the enrichment and differentiation hypotheses of perceptual learning but also, by extension, of the psychophysical as compared to the functional approaches to perception generally.

Lastly, it may be noted that there is a definite connection between the notion of intrinsic and extrinsic cues and the two alternative interpretations of perceptual learning developed on the basis of our criterion. Once a response to a dimension has been established, the possibility for learning on a stimulus-transfer basis, without reinforcement, is dependent on the presence of intrinsic cues which are initially imperfectly differentiated. As for the associative type of learning, it is clearly based on the utilization of extrinsic cues. The postulation of these two types of cues appears, however, to be relevant mainly to discriminations along physical dimensions, while the distinction between the two varieties of perceptual learning may be applied more generally, e.g., to categorization learning (cf. the discussion of Type IV situations, above).

SUMMARY AND CONCLUSIONS

The primary aim of this paper has been to suggest a criterion for perceptual learning which will allow a conceptual distinction between learning based on perceptual functions and that based on response association. The approach suggested is to regard perceptual learning as the development of a generalization of a previously established differential response to a new stimulus. Thus defined, perceptual learning does not require the formation of new S-R associations and, in contrast to associative learning, may occur in the absence of external reinforcement.

---

[5] This shortcoming applies more generally to the experimental work emanating from the empiricist position. Apart from some animal studies, few learning experiments have been carried out by these psychologists, as Gibson and Gibson (16) point out. Apparent exceptions to this statement are the recent studies by Kilpatrick (23) and Weiner (35) on learning in the distorted room, and—representing a somewhat different position—by Kohler (25) on adaptation to the wearing of reversing and distorting lenses. But these are really *unlearning* experiments, in which firmly established responses require modification due to altered stimulus conditions. They can thus not be assumed to represent the process of perceptual learning as it develops originally.

An examination of five types of judgments in the light of this definition brought out a number of specific testable problems with regard to the nature of perceptual learning, but the evidence concerning them is for the most part either unavailable or inadequate on methodological grounds. In several instances, the relevance of the "specificity hypothesis" advanced by Gibson and Gibson (16) was noted. This hypothesis, which describes "what is learned" in perceptual learning as an increased differentiation of variables of stimulation, may be of particular predictive value where the aspects of the stimulus which will be differentiated can be specified. Finally, an experimental approach to the Gibsons' opposition between "enrichment" and "differentiation" theories of perceptual learning was suggested, through the opposition of intrinsic and extrinsic cues in the learning of a discriminative response.

This paper, like that of the Gibsons, has sidestepped the question of the processes and mechanisms within the organism which mediate perceptual learning of the differentiation variety. Admittedly, from a theoretical point of view, the problem of perceptual learning is thus only half solved at best. The task remains to analyze the processes involved in terms of a set of constructs which will permit these phenomena to be integrated into a more general perceptual theory. It is possible that some constructs proposed in other contexts—e.g., Pitts and McCulloch's (29) "scanning mechanisms," or Piaget's (28) "decentrations"—may be adapted to account for improvement in the differentiation of stimuli. At the present, however, they remain largely speculative and of unknown value in the handling of perceptual learning phenomena.

In conclusion, it may be suggested that an analysis of the stimulus, aimed at specifying what aspects are discriminated as a result of learning, may itself contribute heavily to the eventual construction of adequate mediating mechanisms to explain perceptual learning.

REFERENCES

1. ALLPORT, G. W., & PETTIGREW, T. F. Cultural influence on the perception of movement: The trapezoidal illusion among Zulus. *J. abnorm. soc. Psychol.*, 1957, **55**, 104–113.
2. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
3. ARNOULT, M. D. Transfer of pre-differentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401–409.
4. ARNOULT, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 340–351.
5. ATTNEAVE, F. Transfer of experience with a class-schema to identification-learning of patterns and shapes. *J. exp. Psychol.*, 1957, **54**, 81–88.
6. ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.
7. BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.
8. BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.
9. GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.
10. GIBSON, ELEANOR J. Improvement in perceptual judgments as a function of controlled practice or training. *Psychol. Bull.*, 1953, **50**, 401–431.
11. GIBSON, ELEANOR J., & BERGMAN, R. The effect of training on absolute estimation of distance over the ground. *J. exp. Psychol.*, 1954, **48**, 473–482.
12. GIBSON, ELEANOR J., BERGMAN, R., & PURDY, J. The effect of prior training with a scale of distance on absolute and relative judgments of distance over the ground. *J. exp. Psychol.*, 1955, **50**, 97–105.
13. GIBSON, ELEANOR J., & SMITH, JO ANN. The effect of training in distance esti-

mation on the judgment of size-at-a-distance. Human Resources Research Center, Lackland Air Force Base. *Research Bull.,* 1952, No. 52-39.

14. GIBSON, J. J. (Ed.) *Motion picture testing and research.* Washington: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Program Res. Rep. No. 7.)
15. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.
16. GIBSON, J. J., & GIBSON, ELEANOR J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.,* 1955, **62,** 32–41.
17. GIBSON, J. J., & GIBSON, ELEANOR J. What is learned in perceptual learning? A reply to Professor Postman. *Psychol. Rev.,* 1955, **62,** 447–450.
18. HAKE, H. W., & ERIKSEN, C. W. Role of response variables in recognition and identification of complex visual forms. *J. exp. Psychol.,* 1956, **52,** 235–243.
19. HARLOW, H. F. Primate learning. In C. P. Stone (Ed.), *Comparative Psychology.* (3rd ed.) New York: Prentice-Hall, 1951. Pp. 183–238.
20. HASTORF, A. H. The influence of suggestion on the relationship between stimulus size and perceived distance. *J. Psychol.,* 1950, **29,** 195–217.
21. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.,* 1951, **41,** 401–410.
22. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.
23. KILPATRICK, F. P. Two processes in perceptual learning. *J. exp. Psychol.,* 1954, **47,** 362–370.
24. KOEHLER, O. Vorsprachliches Denken und "Zählen" der Vögel. In E. Mayr & E. Schüz (Eds.), *Ornithologie als biologische Wissenschaft.* Heidelberg: C. Winter, 1949. Pp. 125–146.
25. KOHLER, I. Ueber Aufbau und Wandlungen der Wahrnehmungswelt. *Oesterr. Akad. Wiss. Philos.-Histor. Kl.; Sitz.-Ber.,* 1951, **227,** 1–118.
26. NEWMAN, S. E. Effects of contiguity and similarity on the learning of concepts. *J. exp. Psychol.,* 1956, **52,** 349–353.
27. PASTORE, N. Discrimination learning in the canary. *J. comp. physiol. Psychol.,* 1954, **47,** 389–390.
28. PIAGET, J., LAMBERCIER, M., BOESCH, E., & VON ALBERTINI, BARBARA. Introduction à l'étude des perceptions chez l'enfant et analyse d'une illusion relative à la perception visuelle de cercles concentriques (Delboeuf). *Arch. Psychol., Genève,* 1943, **29,** 1–107.
29. PITTS, W., & MCCULLOCH, W. S. How we know universals: The perception of auditory and visual forms. *Bull. math. Biophys.,* 1947, **9,** 127–147.
30. POSTMAN, L. Association theory and perceptual learning. *Psychol. Rev.,* 1955, **62,** 438–446.
31. SEWARD, J. P. The effect of practice on the visual perception of form. *Arch. Psychol.,* 1931, **20,** No. 130.
32. SLACK, C. W. Familiar size as a cue to size in the presence of conflicting cues. *J. exp. Psychol.,* 1956, **52,** 194–198.
33. THORNDIKE, E. L. *Human learning.* New York: Century, 1930.
34. TOCH, H. H., & ITTELSON, W. H. The role of past experience in apparent movement: A reevaluation. *Brit. J. Psychol.,* 1956, **47,** 195–207.
35. WEINER, M. Perceptual development in a distorted room: A phenomenological study. *Psychol. Monogr.,* 1956, **70,** No. 16 (Whole No. 423).

(Received December 30, 1957)

Psychological Review
Vol. 65, No. 5, 1958

THE CONCEPT OF IDENTIFICATION [1]

JEROME KAGAN

Fels Research Institute, Yellow Springs, Ohio

Several years ago Sanford (20) presented an analysis of the concept of identification. In brief, Sanford suggested that the term be applied to situations in which "an individual may be observed to respond to the behavior of other people or objects by initiating in fantasy or reality the same behavior himself . . . the individual strives to behave in a way that is exactly like that of the object" (20, p. 109). Sanford further suggested that the motive for this imitative behavior was a threat to the person's self esteem. By limiting the term "identification" to those imitative behavioral sequences in which the motivation for the act was anxiety over self esteem, Sanford emphasized two points: (*a*) mere similarity in overt behavior between a subject and a model was not necessarily a measure of identification, and (*b*) the motive for the imitative behavior was one of the defining characteristics of an identificatory response.

The various behavioral phenomena which have been labeled "identification" differ in their manifest properties and motivations. The following four classes of behavior have been described as related to the process of identification because they all can lead to similarities in behavior between a subject and a model.

Imitation Learning

This term refers to the initiation and practice of certain responses (gestures, attitudes, speech patterns, dress, etc.) which are not subject to prohibition by the social environment and which are assumed to be the result of an attempt to imitate a model. The behavior has been labeled either "matched-dependent behavior" or "copying" by Miller and Dollard (17). Miller and Dollard posit that initially the imitative act occurs by chance and the act can only be reinforced if some drive is reduced following the execution of the response. According to this view only direct reward from the social environment, like praise or affection, can strengthen the person's tendency to imitate a model. Mowrer (18) distinguishes between developmental and defensive identification. In the former process, the person imitates or reproduces the behavior of a model in order to "reproduce bits of the beloved and longed-for parent" (18, p. 615). Mowrer suggests that most imitation of a model is the result of the desire to reproduce responses which have acquired secondary reward value through association with a nurturant and affectionate model. Thus, Mowrer emphasizes the self-rewarding aspect of certain imitative acts as opposed to Miller and Dollard's emphasis on direct reward from the social environment.

Prohibition Learning

This term refers to the adoption and practice of the prohibitions of the parents and parent substitutes. The acquisition of these prohibitions bears some relation to the process of superego development as described by psychoanalytic theory (2, 3, 4, 11). Several investigators have suggested that a major motivation for the acquisition of some prohibition is anxiety over anticipated loss of love (10, 11, 18, 20, 23). Sanford labeled this process "introjection" and suggested that the learning and maintenance of this class of behavior might be explained without use of the concept of identification.

---

[1] This research was supported, in part, by a research grant (M-1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The views of Wesley Allinsmith, Vaughn J. Crandall, Leonard M. Lansky, and Howard A. Moss are especially acknowledged. A major stimulus for the present essay was a workshop in parent-child relations supported by USPHS Grant 1649 and held at the Merrill-Palmer School, Detroit, Michigan, July 14–27, 1957.

Identification with the Aggressor

This phrase refers to the adoption of behaviors which are similar to those of an aggressive or threatening model. The motivation for this "imitation" is assumed to be anxiety over anticipated aggression or domination by the threatening model. It is difficult to explain this behavior as a product of either prohibition or imitation learning, since the motive and reinforcement do not seem related to anxiety over anticipated loss of love or desire for a direct, social reward like praise or affection. Anna Freud (2) has labeled this phenomenon "identification with the aggressor," Mowrer has called this process "defensive identification" (as distinct from developmental identification), and Sanford has suggested that the term "identification proper" be restricted to this class of behavior.

Vicarious Affective Experience

This phrase refers to the experience of positive or negative affect on the part of a person as a result of an event which has occurred to a model. Salient examples of this phenomenon are (*a*) a child's elation or depression at learning that his parent is a success or failure, or (*b*) a mother's elation following the success of her child in school. This phenomenon of vicarious, affective experience has been attributed to a person's identification with a model, but this affective response has been difficult to explain and often neglected by psychologists investigating the identification process. These four phenomena (imitation learning, prohibition learning, identification with the aggressor, and vicarious, affective experience) appear to be mediated by different motives and rewards, and an analysis of each of them is one purpose of this paper.[2]

In different contexts, social scientists have used the term "identification" to refer to three different sets of variables: (*a*) the process of identification; (*b*) individual differences in the content of the behaviors, motives, and attitudes acquired as a result of the identification process; and (*c*) the differential effect of various models that are used during the identification process (3, 4, 5, 7, 9, 11, 13, 15, 16, 25, 26). This paper recognizes the relevance of the model and content dimensions but is primarily concerned with the process of identification, and will attempt to analyze this process in behavioral terms. It is suggested that the process remains the same regardless of the models used or the specific behavioral content that is acquired as a result of an identification.

Definitions of Identification

The concept of identification originated in psychoanalytic theory, and Freud made a distinction between primary and secondary identification (3, 4, 5). Primary identification referred to the initial, undifferentiated perception of the infant in which an external object was perceived as part of the self, while secondary identification began after the child had discriminated a world of objects separate from the self. Freud implied in his later writings that the process of secondary identification was motivated primarily by the motives and anxieties created by the oedipal situation. In order to reduce the anxiety over anticipated aggression or rejection from the same-sex parent and obtain vicariously the affection of the opposite-sex parent, the child identified with the former. Identification was described by Freud as "the endeavor to mould a person's own ego after the fashion of one that has been taken as a model" (**5**, p. 63).

---

[2] In an unpublished paper presented at a symposium at Harvard University in 1957, Bronfrenbrenner described three types of identification: (*a*) anaclitic identification, (*b*) identification with a source of power, and (*c*) identification through reinforcement of a role model. These three terms are similar in meaning to the present phrases of prohibition learning, identification with the aggressor, and imitation learning, respectively.

Mowrer's concept of "defensive identification," Sanford's definition of "identification proper," and Anna Freud's description of "identification with the aggressor" are all related to the earlier psychoanalytic hypothesis that the threat value of the same-sex parent motivated the child to identify with him in order to reduce the anxiety associated with this threat. However, it is suggested that an individual may identify with a model not only to reduce anxiety over anticipated aggression from a model but also to experience or obtain positive goal states which he perceives that the model commands. The thesis of this paper is that the motivation to command or experience desired goal states of a model is salient in the development and maintenance of an identification.[3] It will be suggested later that two major goal states involved in identification behavior are (*a*) mastery of the environment and (*b*) love and affection. However, it is not implied that these are the only goals which an individual desires to command.

Definition

Identification is defined as an acquired, cognitive response within a person (S). The content of this response is that some of the attributes, motives, characteristics, and affective states of a model (M) are part of S's psychological organization. The major implication of this definition is that the S may react to events occurring to M as if they occurred to him.

The Acquisition and Maintenance of an Identification

Although identification has been defined as a cognitive response, it is not implied that the content of the response is available to consciousness or easily verbalized. Thus the terms "cognitive response," "belief," "wish," or "assumption" will be used in this text to include cognitive processes not always available through verbal report. Identification is not viewed as an all-or-none process. Identification is a response that can vary in strength and there will be differences in the degree to which an S believes that the characteristics of a model, whether assets or liabilities, belong to him. In addition, the S may become identified, to differing degrees, with a variety of models. The motives and reinforcements that are involved in the acquisition and maintenance of this cognitive response are elaborated in the following assumptions.

Assumption 1

Initially the S perceives that the M possesses or commands goals and satisfactions that the S desires. This perception leads to a wish to possess these desired goal states.

Assumption 2

The wish to command the goal states of the M leads to the desire to possess the characteristics of the M because S believes that if he were similar to the M he would command the desired goals. That is, the S assumes that the more similarity there is between the S and M the more likely S is to possess or command the desired goal states of the M.

---

[3] It is assumed that anticipation of a positive goal state is associated with the anticipation of a change in affect, and thus the phrase "experience goal states of the model" will be used synonymously with the phrase "experience affective states of the model." This assumption agrees with McClelland's definition of a motive as an "anticipation of a change in affective state" (**14**, p. 466).

To illustrate, let the S be a child and the M a mother, although S and M could be an adolescent boy and the leader of a group, or a girl and her older sister. The child perceives that the mother can feed the child, restrict the child, obtain articles out of the child's reach, punish the child, etc. Thus, to the S, the M appears to command desired skills and goal states. The discrepancy between the child's perception of his inability to obtain these desired goals and his perception of the more adequate adult elicits the wish to possess or control those goals which he perceives that M commands. The perceptions of the child are subject to distortion, and the child may exaggerate the degree to which M commands desired goals. It was assumed (Assumption 2) that the wish to command these goal states led to the expectation that if S possessed M's characteristics he would also command these desired goals. There often is direct reinforcement of the belief that to "be similar to" a model is equivalent to possessing his positive attributes. Often, the social environment tells the child directly that he is similar to a parent in certain characteristics, and this communication may be contiguous in time with statements related to some of the model's desired goal states. For example, parents and relatives may tell the child "You have your father's eyes," and often add, "You'll grow up to be big and strong just like Daddy." It is suggested that these statements which associate similarities in external attributes with command of desired goal states have an important effect on the child's learning about himself, and lead the child to the expectation that to be similar to the model is equivalent to possessing his positive and desirable attributes.

Assumption 3

The identification response (i.e., "some of the characteristics of the model are mine") is reinforced each time S perceives or is told that he is similar to the M. One type of reinforcement for the identification response occurs when an S is told directly that he and the M are similar in temperament or appearance. It is suggested that a second type of reinforcement for this cognitive response is S's own perception of similarity to the M. Once again, consider the case of the small child and his parent. Although the child may perceive marked differences in size, strength, and skills between himself and the M, he may perceive a similarity in affective states, such as joy, anger, or sadness. The importance of the perception of similarities in affective states between the S and M is stressed because a major motive for identification is a desire to experience positive affective states of the model. Thus, perception of similarity in affect is assumed to have saliency as a reinforcement. If the parent becomes angry, sad, or happy and communicates these affects to the child, the child has the opportunity to perceive that he and the M experience similar feelings. This perception reinforces the belief that there is similarity between the S and M. In addition to similarity in affective states, perception of similarities in external characteristics will reinforce the identification response. With specific reference to the child-parent relation, it is assumed that perception of similarities in sexual anatomy, dress, amount and distribution of hair, and other external attributes are potential reinforcements of the identification. Thus, while the identification response is being learned, the major reinforcements for the response are perceptions of similarity between the S and M.[4] Freud suggested that perceptions of similarity strengthen an identification, for he wrote,

> Identification . . . may arise with every new perception of a common quality shared with some other person who is not an object of the sexual instinct. The more important this common quality is, the more successful may this partial identification become, and it may thus represent the beginning of a new tie (5, p. 65).

Assumption 4

In order for the identification belief to be maintained, the S must not only perceive similarity between the S and M but he also must experience some of the desired, affective goal states of the M. Thus, if the M were successful or happy and S believed that M was experiencing positive affect, the S would also feel positive affect appropriate to the success, and this experience would reinforce his identification. The S also may experience affect appropriate to events occurring to M as a result of the expectation that the social environment will respond to him the same way it responds to the M. That is, when the S has developed some degree of identification with the M he may anticipate that when the social environment praises or rewards the M, it will behave similarly to him. If, on the other hand, the M were sad or criticized, S might experience negative affect because of the identification belief that he and the M were similar and the expectation that the environment might react to him as it did to M. However, if no vicarious command of desired goals or positive affect were experienced as a result of the identification, then the response should extinguish just as any other habit does in the absence of positive reinforcement.[5] That is, some degree of identification should be maintained as long as S perceives that the M commands desired goals. When the S no longer perceives the M in this fashion, then both the motivation for the identification and the intensity of the positive reinforcement should decrease.

The Acquisition of Behavior Similar to a Model: The Motives for Imitation, Identification, and Prohibition Learning

Since perceptions of similarity between the S and M reinforce the identification response, the S may imitate the M during the acquisition phase of an identification in order to increase the degree of similarity. It is acknowledged that the social environment rewards imitative behaviors with affection and praise, and these direct, social reinforcements may strengthen the tendency to imitate adults independently of any identification motives. However, it is suggested, along with Sears et al. (23), that direct, social reinforcement of imitative behavior cannot account for all of the imitative responses that the S initiates. A four-year-old child may simulate adult behaviors when the child is alone or in situations where the parents discourage or punish the imitative response. However, despite the punishment or absence of social reward for some imitative behaviors, the behavior continues to be practiced. Sears et al. call this behavior "role practice" and assume that it is motivated by the "desire to reproduce pleasant experiences" (23, p. 370). Consider the three-year-old girl who plays the role of mother alone in her room. It is hypothesized that a potential reinforcement for this behavior is the creation, in fantasy, of perceptual similarity between the behaviors of the S and M. This perception strengthens S's identification with the M and allows S to share vicariously some of the positive goal states which M commands.

---

[4] It is suggested that the concept of identification has not yielded to a behavioral analysis because the notion of social reinforcement has been viewed as a specific action directed at an individual by a reinforcing agent. There has been a tendency to overlook the possibility that a perception, fantasy, or thought may be a potential reinforcement of a response. A recent experimental finding by Estes and Johns (1) supports the hypothesis that a person's perception of a situation, even though objectively inaccurate, can reinforce his subsequent behavior.

[5] This view of identification suggests a measurement operation which differs from the usual practice of assessing similarities in behavior between an S and an M. One measure of degree of identification would be the degree to which an S's affective state or behavior was influenced as a result of events that occurred to an M. That is, praise or criticism of an M in S's presence should lead to corresponding changes in the affective state of an S who was identified with the M.

A somewhat different phenomenon is the behavior called "identification with the aggressor" by A. Freud or "defensive identification" by Mowrer. Anna Freud describes a girl who was afraid of ghosts and suddenly began to make peculiar gestures as she ran in the dark. She told her brother, "there is no need to be afraid, you just have to pretend that you're the ghost who might meet you" (2, p. 119). The present theory assumes that the child desired the threatening power of the feared object and this motive elicited the imitative behavior. The fantasied perception of similarity to the feared model gave S a vicarious feeling of power and reduced her anxiety over attack. It is suggested that "identification with the aggressor" does not differ from other identification responses with respect to the basic mechanism of acquisition but does involve a specific motive and goal state. Identification with the aggressor involves a specific relationship between the S and M in which S fears the M. Thus, S desires the aggressive power or threat value of the M in order to reduce his own anxiety over anticipated attack. It may be misleading to classify "identification with the aggressor" as qualitatively different from other identificatory behavior merely because the motive and goal differ from those involved in other identifications.

A third motive which can lead to behavioral similarity between an S and M is anxiety over anticipated loss of love or nurturance. It is suggested that many social prohibitions which the M practices are learned by the S in situations in which this anxiety motivate the acquisition and maintenance of the response. The reinforcement for the learned prohibition is continued acceptance and a consequent reduction in anxiety over rejection. The research of Sears et al. (23) suggests a relationship between "high conscience" in a child and a pattern in which the mother is nurturant and uses withdrawal of love as a disciplinary technique. In summary, any one response which is imitative of a model may be mediated by three different motive-reinforcement sequences, and in many instances all three may be involved in producing behavioral similarity between an S and M.[6] Thus, "eating neatly," "getting good grades," or "being nonaggressive" could be motivated by the desire for praise as in imitation learning, by anxiety over loss of love as in prohibition learning, or by the desire to create perceptual similarity between the S and M as in identification. Thus, mere similarity in overt behavior between an S and M may not be the most sensitive measure of degree of identification.

At a more speculative level, it is suggested that the behaviors which have been called "self actualizing" (6) could be motivated and reinforced by a desire for perceptual similarity to an M and be an indication of early identification tendencies. Even the most orthodox supporters of the importance of simple imitation learning find it difficult to explain the child's initial imitations of a model. Once the child has begun to imitate a model it is likely that praise and recognition from adults could maintain this behavior. However, why does the child suddenly want to dress himself, sit on the toilet alone, or put on Daddy's shoes? It is difficult to account for the initial display of this imitative behavior, and the term "self actualization" implies that the child has some biological drive to use his potentialities. This hypothesis seems no more parsimonious than the suggestion that the initiation of these "self actualizing" behaviors is motivated by S's desire to create perceptual similarity between himself and a model.

---

[6] In a manuscript being prepared for publication, H. Kelman suggests that the response of conformity to the attitudes of another person can be mediated by three different motives. His analysis of conformity parallels the present discussion of imitative behavior.

Two Goals Motivating Identification: Mastery and Love

It has been assumed that S's desire to command certain goal states motivates his identification with a model. It is suggested, for the child especially, that two important goal states that the S desires to command are (*a*) a feeling of power or mastery over the environment and (*b*) love and affection. Attainment of these goals should lead to diminution in anxiety over helplessness or loneliness. The young child perceives that he is not able to gratify all of his needs while the parental model is perceived as more capable of dealing with the environment. This discrepancy between the S's perception of his own relative helplessness and the power that he perceives M to possess motivates the wish to have M's power and the search for perceptions of similarity between himself and the M.

Unfortunately, there are no empirical studies which directly test these hypotheses because most of the research on identification has used similarities in behavior between an S and M as the measure of identification. However, there are some results which are at least consistent with the view that the child identifies with the more powerful parent and the one who is perceived to command important sources of gratification. Payne and Mussen (19) reported that adolescent boys who perceived the father as rewarding on projective tests were more highly identified with the father (based on similar answers to a personality inventory) than boys who pictured their fathers as nonrewarding. In addition, boys with dominant and "masculine" mothers tended to be poorly identified with the father. P. S. Sears (22) reported a finding that is more difficult to explain without use of the concept of identification. She found, in a doll-play situation, that kindergarten girls used the mother doll as agent significantly more often than the father doll, while boys used both mother and father dolls with more nearly equal frequency. Since the mother is initially the major controller of gratifications for both sexes, one might expect an initial identification with her for both boys and girls. P. S. Sears (22) also reports that the kindergarten boys who used the mother doll most often had mothers who were (*a*) more nurturant than the father, (*b*) more critical of the father, and (*c*) more restrictive of the child's mobility outside the home. This result is consistent with the hypothesis that the child is predisposed to identify with the parental model who is perceived as controlling important goal states.

A study of Maccoby and Wilson (15) furnishes more direct support for the present hypotheses. The authors showed movies to seventh grade boys and girls and then determined the protagonist with whom the child identified. The most significant result was that a "boy's choice of screen character (the one with whom he was presumed to identify) is more closely related to the social class level *to which he aspires* than to the level his family currently occupies" (15, p. 79). This result suggests that the child identified with models who commanded desired goals.

A second goal state which may motivate identification is the desire for nurturance and affection. In addition to Freud's classical hypothesis that the child identified with the same-sex parent in order to receive vicariously the affection of the opposite-sex parent, there are situations in which nonparental models command sources of affection. The relation between siblings is such a situation, and the younger child may identify with an older sibling if the former perceives that the latter commands parental affection. The research of Helen Koch (12) indirectly supports this hypothesis. She reported that school-age boys with older sisters tended to develop more feminine attributes than boys with older brothers. On the other hand, girls with older brothers tend to be more masculine than girls with older sisters. In the experiment of Maccoby and Wilson, described earlier, the authors reported that girls were more likely than boys to recall movie content involving boy-girl interaction while boys were superior on recall of aggressive acts by the hero. If one assumes that the need for affection is stronger for girls than for boys, and that the recalled content is influenced by the model chosen for identification, then these results suggest that the specific goal states desired by the S determine the models chosen for identification.

Factors Influencing the Strength of Identification

The strength of the identification habit, following a basic behavioral law, should be a function of the strength of the motive and the quality and frequency of the reinforcement (8). It would be predicted, therefore, that the most intense identification would occur when the S had strong needs for love and power, felt incapable of gratifying these motives through his own skills, and perceived similarity between himself and an M who commanded these goals. Utilizing this hypothesis, two generalized predictions can be made concerning the strength of identification for different ages and models.

1. The strength of identification tendencies should decrease with age because, in general, the individual's ability to gratify his needs for mastery and love through his own behavior, rather than through a vicarious mechanism, should increase with development. Thus, the identifications of a young child should be more intense than the identifications of older individuals.

2. An identification with an M with whom S was in direct contact should be stronger than with an M with whom S was not in contact, assuming that the motivation for identification was constant and the models were perceived as equally potent. This statement is based on the assumption that the reinforcements of perceived similarity are stronger when S perceives the affects and attributes of the M directly as opposed to instances in which he is merely told that he is similar to the M. Thus, degree of identification with a father with whom S was in contact should be greater than with an imagined fantasy father whom S had never seen. Only very indirect evidence is available to support this prediction. However, reports by P. S. Sears (21) and Sears

et al. (24) suggest that absence of the father from the home tends to decrease the degree of "masculine" doll play in preschool boys while this experience has little effect on the doll play of girls. The results are open to alternative interpretations but are not inconsistent with the present hypothesis.

SUMMARY

This paper has attempted to analyze the concept of identification and place the concept within a learning-theory framework. Identification was defined as an acquired, cognitive response. The content of this response was that some of the characteristics of a model belonged to the individual and the individual behaved as if some of the characteristics and affective states of the model belonged to him. Identification was not viewed as an all-or-none process. An identification can vary in strength and the individual can identify, to differing degrees, with a variety of models. The motive for the acquisition and maintenance of the identification response was a desire for the positive goal states commanded by the model, and mastery of the environment and love-nurturance were suggested as two important goals. The reinforcement for the acquisition of the identification was perceived similarity in attributes between the person and the model. Thus, the person may strive to imitate aspects of the model's behavior in order to create perceptual similarity between himself and the model. Once the identification was established, the individual behaved as if the goal states of the model belonged to him and the positive affect derived from this vicarious sharing of desired goal states helped to maintain the identification.

It was suggested that the usual emphasis on similarities in overt behavior between an individual and a model is not the best measure of identification, since the motives and reinforcements involved in imitation and prohibition learning could also explain similarities in behavior between two people. A differentiation of imitative behavior based on imitation learning, prohibition learning, and identification was attempted.

REFERENCES

1. ESTES, W. K., & JOHNS, MARCIA D. Probability learning with ambiguity in the reinforcing stimulus. *Amer. J. Psychol.*, 1958, **71**, 219–228.
2. FREUD, ANNA. *The ego and the mechanisms of defense.* London: Hogarth, 1937.
3. FREUD, S. *New introductory lectures in psychoanalysis.* New York: Norton, 1933.
4. FREUD, S. *The ego and the id.* London: Hogarth, 1935.
5. FREUD, S. *Group psychology and the analysis of the ego.* London: Hogarth, 1949.
6. GOLDSTEIN, K. *The organism.* New York: American Book, 1939.
7. GRAY, SUSAN W., & KLAUS, R. The assessment of parental identification. *Genet. Psychol. Monogr.*, 1956, **54**, 87–114.
8. HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
9. KAGAN, J. The child's perception of the parent. *J. abnorm. soc. Psychol.*, 1956, **53**, 257–258.
10. KAGAN, J. Socialization of aggression and the perception of parents in fantasy. *Child Develpm.*, 1958, **29**, 311–320.
11. KNIGHT, R. P. Introjection, projection and identification. *Psychoanal. Quart.*, 1940, **9**, 334–341.
12. KOCH, HELEN L. Attitudes of young children toward their peers as related to certain characteristics of their siblings. *Psychol. Monogr.*, 1956, **70**, No. 19 (Whole No. 426).
13. LAZOWICK, L. M. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51**, 175–183.
14. MCCLELLAND, D. C. *Personality.* New York: Sloane, 1951.
15. MACCOBY, ELEANOR E., & WILSON, W. C. Identification and observational learning from films. *J. abnorm. soc. Psychol.*, 1957, **55**, 76–87.
16. MARTIN, W. E. Learning theory and identification: III. The development of

value in children. *J. genet. Psychol.*, 1954, **84**, 211–217.

17. MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

18. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

19. PAYNE, D. E., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, **52**, 358–362.

20. SANFORD, R. N. The dynamics of identification. *Psychol. Rev.*, 1955, **62**, 106–118.

21. SEARS, PAULINE S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**, No. 6 (Whole No. 323).

22. SEARS, PAULINE S. Child rearing factors related to playing sex-typed roles. *Amer. Psychologist*, 1953, **8**, 431. (Abstract)

23. SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

24. SEARS, R. R., PINTLER, MARGARET H., & SEARS, PAULINE S. Effect of father separation on pre-school children's doll play aggression. *Child Develpm.*, 1946, **17**, 219–243.

25. SEWARD, J. P. Learning theory and identification: II. The role of punishment. *J. genet. Psychol.*, 1954, **84**, 201–210.

26. STOKE, S. M. An inquiry into the concept of identification. *J. genet Psychol.*, 1950, **76**, 163–189.

(Received February 27, 1958)

Psychological Review
Vol. 65, No. 5, 1958

INTRALIST GENERALIZATION IN PAIRED-ASSOCIATE LEARNING[1]

BENNET B. MURDOCK, JR.

University of Vermont

It seems to be generally accepted that, during the course of paired-associate learning, generalization tendencies increase to a maximum and then decrease. This hypothesis is endorsed by Osgood (16, pp. 537–539), and stated without comment by Bugelski (3, p. 410), Hovland (12, p. 666), and Underwood (17, p. 408). This increase and decrease in generalization tendencies during verbal learning was first suggested by Gibson (8), who made it an integral part of her theoretical formulation. Her Postulate 4 states, "Generalization will increase to a maximum or peak during the early stages of practice with a list, after which it will decrease as practice is continued" (8, p. 206). In this article we would like to suggest an alternate but opposing hypothesis, that generalization is greatest at the start of paired-associate learning and only decreases with practice.

In a paired-associate list *n* pairs long, four different types of responses can be distinguished: correct responses (S_a—R_a, S_b—R_b,... S_n—R_n); incorrect responses (S_a—R_b, S_a—R_c,... S_a—R_n; S_b—R_a, S_b—R_c,... S_b—R_n, etc.); intrusions (any responses other than R_a, R_b,... R_n); and omissions (no response). Generalization is ". . . the tendency for a response R_a learned to S_a to occur when S_b (with which it has not been previously associated) is presented" (8, p. 204). According to this definition, an incorrect response would be a generalized response, as both are essentially the right response to the wrong stimulus. An intrusion would not be a generalized response to any stimulus in the list because it is not a correct response to any stimulus in the list. And obviously correct responses and omissions are not generalized responses. Therefore, it would seem (and this is the position taken here) that in paired-associate learning incorrect responses and generalized responses are one and the same.[2]

On a common-sense level it seems unreasonable to consider errors early in learning as generalized responses. If the correct response has not been learned (i.e., if *S* has not yet learned S_a—R_a), how could an incorrect response indicate generalization? Also, the definition of generalization given above states that the response R_a must be learned to the stimulus S_a before S_b—R_a could be considered a generalized response. However, if the response R_a does in fact occur to stimulus S_b, this would almost of itself indicate that some learning had already occurred; the probability of the occurrence of this response (or for that matter any other response on the list) by chance is so slight as to be negligible. Therefore, it would seem quite unlikely that an incorrect response could occur before some learning had taken place.

Equating incorrect and generalized responses is desirable from the point of view of simplicity. The definition of a generalized response in paired-associate learning becomes clear cut and unambiguous when the two are considered synonymous. There is no need to set up some arbitrary point in the learning before which incorrect responses would not be generalized responses but after which incorrect responses would be generalized responses. And, as will be pointed out in the final section of this paper, this position results in a prediction about transfer of training which is apparently substantiated by the available data. Therefore, it is felt that incorrect responses at any stage in paired-associate learning should be considered generalized responses.

---

[1] This work was supported by a grant, G-2590, from the National Science Foundation. The author would like to thank Spaulding Rogers, Norman J. Slamecka, and Ernest N. Damianopoulos for a critical reading of the manuscript.

[2] An incorrect response could also indicate response generalization. However, in this article it is assumed that the response terms are sufficiently dissimilar and/or predifferentiated so that no response generalization occurs; all generalization will therefore be considered stimulus generalization.

In the standard paired-associate learning task *S* is not informed as to what the responses are before he actually starts to learn the list. Therefore, he must learn that the responses are R_a, $R_b, \ldots R_n$. This may involve learning to make responses that have never previously been made, and it will certainly involve learning to restrict the range of responses to R_a, $R_b, \ldots R_n$. Also, *S* must learn the pairings. That is, he must learn S_a—R_a, S_b—$R_b, \ldots S_n$—R_n. Furthermore, within limits he is free to respond or not as he wishes. Therefore, on the first trial of a paired-associate learning task which uses the anticipation method, the probability of an incorrect or generalized response is very slight. If *S* learns to a criterion of so many perfect trials, the probability of an incorrect or generalized response at the end of learning is also very slight. Assuming that some generalization occurs, there obviously must be first an increase, then a decrease in the frequency of incorrect or generalized responses. This does not mean, however, that there is an increase and then a decrease in generalization, for under these conditions the frequency of incorrect or generalized responses is not an appropriate measure of generalization.

The frequency of incorrect responses is an inappropriate measure of generalization because omissions can occur, and they introduce a systematic bias. Omissions are most frequent at the start of learning, yet this is exactly the time when the frequency of generalized responses is low. If omissions were prohibited and *S* forced to respond to each stimulus, almost certainly most of the responses (excluding for the moment intrusions) would be incorrect or generalized responses, not correct responses. Therefore, contrary to what Gibson assumed (9, p. 194), the frequency of overt generalized responses is not a satisfactory measure of generalization tendencies when omissions are possible, for omissions result in a serious underestimation of the amount of generalization that would occur were *S* forced to respond.

Intrusions introduce a further complication; they too, if allowed to occur, would be most frequent early in learning. In the standard paired-associate procedure *S* must learn both the responses and the pairings, and intrusions indicate that *S* has not yet learned the responses perfectly. But, as Hovland and Kurtz (13) have pointed out, this is a separate problem from the learning of the pairings (or, in the case they discuss, the order in a serial list). If the chief purpose is to study generalization, the responses are merely a means to an end; one draws conclusions about confusion among the stimuli from the responses they evoke. Therefore, learning the responses merely introduces a complication that should be avoided.

It is felt, then, that to study generalization tendencies during learning both omissions and intrusions should be ruled out. This would not be particularly difficult to do. To exclude omissions, forced anticipation could be used;

S would be instructed always to respond during the "stimulus-only" period, even if the response were a pure guess. To exclude intrusions, *S* would be told the responses (though naturally not the pairings) in the instructions. If the responses were unfamiliar, a familiarization procedure such as Hovland and Kurtz (13) used could precede the learning. Under these conditions the frequency of incorrect or generalized responses would be an adequate measure of generalization tendencies, since omissions and intrusions had been ruled out.

Under these conditions only two types of responses would be possible, correct responses and incorrect or generalized responses. Assuming that *S* followed instructions and assuming that learning occurred, the frequency of correct responses would (disregarding inversions and plateaus) increase regularly with practice. Concomitantly, the frequency of incorrect or generalized responses would decrease regularly with practice. And, since it is felt that these are the appropriate conditions under which to measure generalization, it would seem that generalization tendencies have to be at a maximum at the start of learning and decrease with practice. There is, therefore, only a decrease, and no increase, in generalization during paired-associate learning.

The foregoing has been based on the assumption that the similarity among stimuli is the same. That is, no two stimuli are more similar than any other two. This is probably impossible to achieve in practice, and is often deliberately violated in order to get groups of stimuli within the total list with high intragroup similarity. If S_a and S_b are very similar to one another, the probability of S_b—R_a may be greater after a certain amount of practice than it was originally. However, the fact that the probability of this one specific generalized response has increased does not mean that the probability of all generalized responses to S_b has increased. At the start of learning (assuming no omissions or intrusions), the probability that S_b—R_b will occur by chance would be $1/n$, the probability that S_b—R_a will occur by chance would also be $1/n$, and the sum of the probabilities that other incorrect or generalized responses will occur by chance would be $(n-2)/n$. Even though the probability of S_b—R_a may have increased with practice (and it may increase appreciably), provided that the probability of the correct response S_b—R_b does not drop below chance (i.e., $1/n$), then this increase must be accompanied by a corresponding decrease in the probability of other incorrect or generalized responses (i.e. $(n-2)/n$). Therefore, under these conditions there can be no net increase in generalization.

Suppose, however, that the probability of the correct response S_b—R_b did actually drop below chance $(1/n)$. This could conceivably occur if S_a—R_a was learned (i.e., if the probability of S_a—R_a increased from $1/n$ to some value X, where $1/n < X \leqslant 1.00$), if S_b—R_b was not learned, and if S_a and S_b were sufficiently similar (in the extreme case indistinguishable) to produce generalization. Assume that there is a decrease in the probability of S_b—R_b; the maximum decrease possible would be from $1/n$ to 0, or $1/n$. At the same time there must, of necessity, be an increase in the probability of a correct response to S_a. This increase would be from $1/n$ to X, or $X - 1/n$. The only way in which there could be a net increase in generalization tendencies would be for the decrease in the probability S_b—R_b to be greater than the increase in the probability S_a—R_a; i.e., for $1/n$ to be greater than $X - 1/n$. For this to occur, X would have to be smaller than $2/n$, or the increase in the probability of the correct response

S_a—R_a would have to be less than $1/n$. With a paired-associate list of any appreciable length this maximum allowable increase in probability would be so slight as to make it doubtful that any generalization could occur after so little learning. So the conditions necessary for a net increase in generalization would seem to rule out the possibility that any generalization will occur.

It is possible, of course, for more than two stimuli to be similar. If two or more stimuli are similar to S_a, the maximum allowable increase in the probability of S_a—R_a becomes n'/n, where n' is the number of stimuli similar to S_a. The net increase in generalization then becomes $n'/n - (X - 1/n)$, or $(n' + 1)/n - X$. This is the most general expression, and, as before, the smaller X is, the more generalization tendencies could increase but the less likely it is that any generalization could occur. However, as X approaches 1.00, a net increase in generalization becomes impossible no matter how many stimuli are similar to S_a. So, to summarize, because the increase in the probability of one specific generalized response must be accompanied either by a decrease in the probability of other generalized responses to the same stimulus or by the increase in the probability of a correct response to a similar stimulus, it seems highly unlikely that, even with unequal intralist stimulus similarity, there ever could be a measurable increase in generalization tendencies with practice.

The hypothesis presented here suggests that there is only a decrease in generalization tendencies, yet studies of conditioning have shown an initial increase. One explanation for this difference is that in paired-associate learning all stimuli are presented to S from the first trial on. Learning requires the formation of discriminations, and generalized responses must be extinguished if a criterion of learning is to be reached.

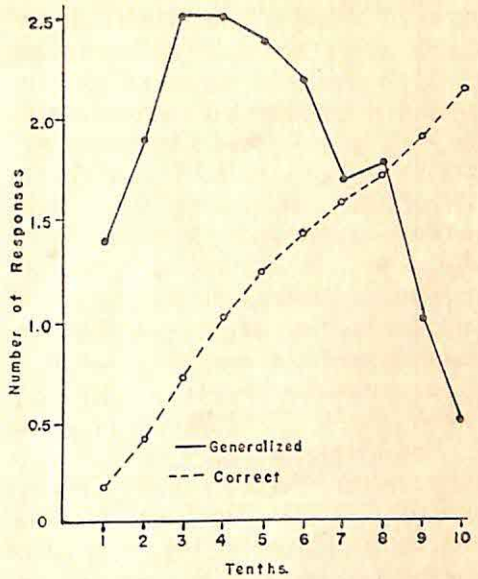


FIG. 1. Curve for number of correct and generalized responses during learning (copied from Gibson's Fig. 3 [9, p. 195]).

On the other hand, in studies of classical conditioning the stimuli used in the test for generalization are customarily not presented until the conditioned response has become established, and there is no opportunity for prior extinction of generalized responses.

RELATED STUDIES

At least two studies, one by Gibson (9) and one by Gagné (4), have reported an increase and then a decrease in generalized responses in a standard paired-associate learning task. The procedures of the two experiments were very similar, and apparently omissions and intrusions were possible in both. Therefore, the results of these experiments do not prove that generalization increased before decreasing. However, even allowing for the procedure used, the evidence they present for an initial increase in generalization is less impressive than first appears.

In the Gibson study, Vincent curves were plotted for correct and generalized

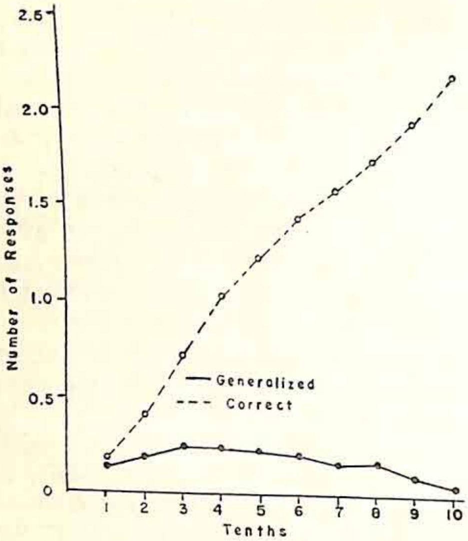


FIG. 2. Fig. 1 redrawn with curve for number of generalized responses to scale.

responses. In her Fig. 3, redrawn here as Fig. 1, the curve for generalized responses appears to do exactly what it is supposed to do, increase to a maximum early in learning and then decrease. However, this is misleading because the two curves are not drawn to the same scale. The scale of one is ten times that of the other. When drawn to the same scale (see Fig. 2) the increase in generalized responses is relatively trivial compared to the increase in the frequency of correct responses that is occurring at the same time.

The reason for emphasizing changes in generalized responses in relation to changes in correct responses is that a prior study by Hovland (11) dealt with the relative strength of a generalized response throughout learning. This was a study of generalization as a function of the number of reinforcements in classical conditioning, and was evidently influential in suggesting to Gibson the analogous hypothesis for verbal learning. Hovland measured the strength of a conditioned GSR after 8, 16, 24, and 48 paired presentations. Actually, the absolute strength of the generalized response increased with the number of reinforcements, but as it increased less at the higher levels than the strength of the conditioned response, there was eventually a relative decrease in generalization.

In the Gagné study (4) the errors also increased to a maximum and then decreased. Table 1 shows the number of correct responses, the number of generalized responses, and the ratio of correct to generalized responses for Group 1, the group that demonstrated the greatest effect. As the ratio shows, there was a slight initial increase in generalized responses. However, the increase was only from 3.11 to 3.27 while the decrease was far larger, from 3.27 to 0.22. So here, as in the Gibson study cited above, on a relative basis the increase was not impressive. And, it is felt, had omissions and intrusions been ruled out, there would not have been any increase at all in generalized responses.

TABLE 1

TRIAL-BY-TRIAL RECORD OF NUMBER OF GENERALIZED RESPONSES, CORRECT RESPONSES, AND RATIO OF GENERALIZED TO CORRECT RESPONSES[a]

Trial	Generalized Responses	Correct Responses	Ratio
1	2.27	0.73	3.11
2	3.87	1.20	3.22
3	4.80	1.47	3.27
4	4.93	2.33	2.12
5	4.40	3.80	1.16
6	4.73	3.53	1.34
7	4.73	4.20	1.13
8	4.13	4.87	0.85
9	3.67	5.40	0.68
10	3.80	6.00	0.63
11	2.67	6.20	0.43
12	2.47	7.20	0.34
13	1.73	7.67	0.23
14	1.80	8.20	0.22

[a] For Group 1 from Table 1 and Table 2 of Gagné (4, pp. 65 and 67).

In addition to the experiments of Gibson and Gagné, there is an experiment by Gagné and Foster (6) which also appears to confirm the Gibson hypothesis. This was a choice reaction-time experiment in which *S*s were given 0, 10, 30, or 50 trials on two of the four components of the final task prior to 60 trials on the complete task itself. Gagné and Foster used the number of errors as a measure of generalization, and reported two main findings which appear to support Gibson. First, the total number of errors on Task 2 increased and then decreased as a function of number of trials on Task 1. That is, the 10-trial group made more errors, but the 30-trial and 50-trial group made fewer and many fewer errors than the 0-trial (control) group, respectively. This increase and decrease in the total number of errors is shown clearly in their Fig. 4 (6, p. 60). The second finding in support of Gibson is that the point of maximum errors (i.e., maximum generalization) appears to shift to the left as practice on Task 1 increases. As they report,

> . . . the point of maximum generalization [abbreviated PMG] seems to be passed in the zero-trial group at about the 40th trial; in the 10-trial group at about the 30th trial; in the 30-trial group at about the 15th trial. The group having 50 trials of preliminary training does not exhibit any marked drop in error score (6, pp. 60–61).

Not only do generalization tendencies increase, according to Gibson, early in learning but also—and this is what Gagné and Foster seem to overlook—after the PMG they should decrease. Assuming comparable groups, assuming that training on Task 1 makes the PMG occur earlier on Task 2, and assuming that the more training on Task 1 the earlier the PMG, then according to Gibson's theory a group with more training on Task 1 should make more errors on Task 2 before its PMG than the other group. However, after the PMG of the second group, the first group should make fewer errors. For instance, the 0-trial group reached its PMG at 40 trials and the 10-trial group reached its PMG at 30 trials. Before 30 trials the 10-trial group should make more errors than the 0-trial group. However, after 40 trials this should reverse; the 0-trial group should make more errors than the 10-trial group. (The relative frequency of errors between the two PMGs cannot be predicted without knowing the exact shape of the curves.)

If this reasoning is correct, then it is inappropriate to compare groups in terms of total number of errors made in Task 2. Instead, it should be predicted that the group with the earlier PMG should make more errors early but fewer errors late in Task-2 learning. Since Gagné and Foster report the number of errors for each group in blocks of ten trials each, and since they estimate the PMG of three groups, it is possible to use their data to test Gibson's theory in this way. Each of the three groups was compared with the other two, and the direction of the predicted difference in errors for each comparison is shown in Table 2. In each case the prediction was based solely on the relative position of the two PMGs, and no predictions were made for the trials falling between the two PMGs. The actual results were determined from Fig. 3 (6, p. 59) in Gagné and Foster's article. Each prediction was considered either correct or incorrect, and this is shown in Table 2 by plus and minus signs. As can be seen, the predictions based on Gibson's theory are correct seven times and incorrect five times. Thus, the theory would not seem to be appreciably better than chance. So here, as in the other experiments, the results fail to provide clear-cut proof that generaliza-

TABLE 2

Predictions as to Which of Two Groups Should Show More Errors in Each Block of Trials[a]

Comparisons	Blocks of Ten Trials					
	1–10	11–20	21–30	31–40	41–50	51–60
0-trial with 10-trial	$0<10(+)$	$0<10(+)$	$0<10(+)$	—	$0>10(-)$	$0>10(-)$
0-trial with 30-trial	$0<30(+)$	—	—	—	$0>30(-)$	$0>30(-)$
10-trial with 30-trial	$10<30(-)$	—	—	$10>30(+)$	$10>30(+)$	$10>30(+)$

[a] A dash indicates that no prediction can be made.

tion tendencies increase to a maximum and then decrease.[3]

Predictions

The question of whether or not generalization tendencies increase early in learning is of some importance in that the two points of view result in different predictions. In a transfer situation with an A—B, A—C paradigm, *S* is learning a set of responses to old stimuli. If, as Gibson says, a low level of practice on A—B results in an increase in generalization, then it should follow that negative transfer would occur; a little practice would be worse than none at all. On the other hand, if generalization only decreased there should be no negative transfer under any conditions, only positive transfer increasing as the amount of practice of the first task increased. Gibson discusses this case in her Proposition XV, "Learning of Pre-differentiated Items" (8, p. 222), where she points out that positive transfer should occur in the learning of well-differentiated items.

Studies in verbal learning using an A—B, A—C paradigm cannot be used to test this prediction because the interlist response interference may obscure any effects due to predifferentiation. However, the so-called "stimulus predifferentiation" studies (1, 14) would appear to provide an ideal test for these predictions. The typical predifferentiation study uses an A—B, A—C paradigm with a relatively high level of intralist stimulus similarity, thus maximizing the possibilities for reduction of generalization tendencies. Low interlist response similarity is also common, and this minimizes the possibilities of associative inhibition. Thus, in Task 1 *S* may have to learn names for complex visual stimuli, while in Task 2 *S* has to learn simple motor discriminative responses to the same stimuli (15). According to Gibson, as degree of Task 1 learning increases, negative transfer should increase to a maximum, then decrease, giving way to positive transfer with sufficient practice on Task 1.

Recently Arnoult has made a survey of predifferentiation studies, and concludes that there is "enough agreement in results among various experiments to provide generalizable conclusions in two broad areas" (1, p. 339). Fortunately, one of these two areas is the amount of predifferentiation training, and his generalization is, *"Positive transfer from stimulus predifferentiation training may*

---

[3] Three other experiments (2, 5, 7) in this series of studies varied the degree of first-task learning so as to permit a test of the Gibson hypothesis. In not one of these three experiments was there any indication that a low level of first-task learning resulted in an increase in errors on the second task. Actually the reverse was generally true; in all three experiments the more practice on the first task, the fewer the errors that usually occurred early in the second-task learning.

be expected after a minimum of 4 to 8 pretraining trials and reaches a maximum after 8 to 12 pretraining trials" (1, p. 341). Further on in the same article Arnoult discusses the specific prediction resulting from Gibson's hypothesis, and states (1, p. 344) that there is no evidence ("negative results") for negative transfer. Thus, the studies reported to date give no support to Gibson's position. On the other hand, they are consistent with the viewpoint that generalization is greatest at the start of learning.[4]

SUMMARY

Gibson has suggested that during the course of paired-associate learning generalization increases to a maximum and then decreases. It is suggested here that generalization is maximum at the start of paired-associate learning and only decreases with practice. Changes in generalization cannot be inferred from changes in overt errors in the typical paired-associate procedure, for omissions and intrusions can occur, and these introduce a systematic bias. When the procedure is such that neither omissions nor intrusions can occur, the frequency of incorrect responses does provide a siutable measure of generalization, and under these conditions if learning occurs generalization can only decrease with practice.

One prediction that results from the position presented here is that with an A—B, A—C transfer paradigm and low interlist response interference, positive transfer should increase as the degree of first-task learning increases. Studies of stimulus predifferentiation provide an appropriate test, and the results of these studies appear to support this prediction.

REFERENCES

1. ARNOULT, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.
2. BAKER, KATHERINE E., & WYLIE, RUTH C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632–638.
3. BUGELSKI, B. R. *The psychology of learning*. New York: Holt, 1956.
4. GAGNÉ, R. M. The effects of sequence of presentation of similar items on the learning of paired-associates. *J. exp. Psychol.*, 1950, **40**, 61–73.
5. GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
6. GAGNÉ, R. M., & FOSTER, HARRIET J. Transfer of training from practice on components in a motor skill. *J. exp. Psychol.*, 1949, **39**, 47–68.
7. GAGNÉ, R. M., & FOSTER, HARRIET J. Transfer to a motor skill from practice on a pictured representation. *J. exp. Psychol.*, 1949, **39**, 342–354.
8. GIBSON, ELEANOR J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
9. GIBSON, ELEANOR J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.
10. GOSS, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.
11. HOVLAND, C. I. The generalization of conditioned responses. IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, **21**, 261–276.
12. HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
13. HOVLAND, C. I., & KURTZ, K. H. Experimental studies in rote-learning theory:

---

[4] Goss (**10**) incorporates the Gibson concept of increase and decrease in generalization into his theoretical system, and makes the prediction that high similarity stimuli will actually result in less transfer than stimuli of medium similarity. Here too the prediction from Gibson's theory seems to be wrong; in a recent study (**15**) in which similarity was varied, the high-similarity group was the only group showing significant positive transfer.

X. Pre-learning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

14. MURDOCK, B. B., JR. Transfer designs and formulas. *Psychol. Bull.*, 1957, **54**, 313–326.

15. MURDOCK, B. B., JR. The effects of task difficulty, stimulus similarity, and type of response on stimulus predifferentiation. *J. exp. Psychol.*, 1958, **55**, 167–172.

16. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

17. UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

(Received March 14, 1958)

Psychological Review
Vol. 65, No. 5, 1958

TOWARD A UNITARY THEORY OF PERCEPTION

KAI VON FIEANDT

University of Helsinki

Perceptual psychology today inclines toward phenomenological terminology. It deals chiefly with impressions of experienced reality. Recently perceptionists have been interested in physiological correlates of these impressions. Because very complicated physiological relations underlie even the primary or simple perceptual processes, is it not natural, then, that psychological terminology should be tentatively phenomenological? Only consider a few of the contents of immediate experience that can be easily described in phenomenological terms—round, complex, closed, threatening, balanced, precise, adaptive—which are functionally dependent on very complicated underlying structures. In the words of Eino Kaila, "Considered functionally the Gestalt quality is always very complicated: actually it can be simply described" (4, 5). Also from modern personality psychology one can reach the same conclusion. Personality diagnosticians employ descriptive (phenomenological) expressions from colloquial speech which often represent rich unanalyzable manifolds of functional relationships.

The problem of a unified explanation of our perceptual world implies a research program to investigate the conditions of our impressions of objects. Why do we assume a "substance" or a "material" as the carrier of perceptual thing-quality? The phenomenal object gains in reality when the impression is multidimensional. The material in question must not only offer itself optically, but also preferably tangibly. We adopt in this respect the classical reality criteria of St. Thomas. By means of multidimensional analysis we can, in my opinion, also achieve scientific progress. However phenomenological the terminology, we can probably discern certain systems of relations by means of an analysis of our coordinated sensations, thus avoiding the diffuse impression of "substance." When proceeding to map these systems of configurations in terms of quantitative relations, the science of psychology can make progress towards becoming a natural science.

In my opinion, the most important contribution to this problem was made by David Katz in *The Modes of Appearance of Colours and Their Dependence on Individual Experience,* 1911, later revised as *The World of Colour,* 1930 (7, 8). When he described the phenomena of "surface color," "film color," and "space color," he could not guess how important these ventures would later prove to be for psychological theory. When he referred to the meaning of the reduction-screen and formulated his principle of "field-size," he implied a setting rich in the relational theory of explanation, to which we today are indebted for the possibility of intermodal theory. We must give Katz the credit he deserves when he maintains, in his *Gestalt Psychology* (9), that we previously were concerned too exclusively about the psychology of the visual sense. Probably there is for this preference a perceptual-psychological explanation. There exists in the optical realm the possibility of combining certain cues for "thingness" with more elementary color qualities—for example, microstructure of the surfaces, form and size relations, general lighting, etc. The perceived colors can, in other words, be exclusively related to optical phenomena. The conditions of object-perception can consequently be reduced by operations not exceeding the optical sphere. And vice-versa: one can objectify "pure" color sensations without leaving the optical modality; one simply coordinates them with form impressions. The leading position of the visual sense makes this many-sidedness possible. There is no reason why color impressions should immediately be referred to other modal contents, for optical form and space are best adapted to "take them over."

Similar considerations apply to the perceived world of touch. In his *Structure of the Touch-World* (7), Katz describes the modes of appearance of tactual experience. How is this modality objectively rendered (objectivized) in terms of palpable images? Only men with normal visual images (not the born-blind) are able to objectivize their exclusively tactual impressions. Their optically constructed coordinate-system is obviously able to supply information about simultaneous shape-gestalts in the environment.

Let us consider some examples:

Surface-touch provides us a thing-impression whatever the position and the distance of the touched object.

Space-filling touch will be directly experienced in walking in a strong wind. A similar experience is also felt when one moves one's extremities quickly through a liquid. These cues hardly provide a basis for objectivizing, for a definitive object cannot be experienced. Therefore one could consider this space-filling phenomenon as an analogy of film color.

Space-filling surface quality can bring with it objectification, especially as it concerns itself with a relatively small hard object behind the material in question. The situation cannot be compared with the perception of a surface color "through" or "behind" colored illumination. The configuration brought out by illumination has its counterpart in the working of the tactual (kinematic) or haptic framework. The movement of the hands or of the felt surface yields an articulation comparable to the effect of illumination in vision.

The born-blind reach a certain state of phenomenal objectivity on the basis of their simultaneous haptic and kinesthetic impressions. But, as is well known, their tactile-kinesthetic gestalten remain only formless and diffused (12).

The sense of hearing can mediate thing-impressions only if the auditory stimulation comes combined with optic or haptic stimuli simultaneously and is localized in the same direction. The tendency to identify auditory impressions with visible things is shown in spiritualistic seances when ventriloquist mediums create illusions of sounds issuing from megaphones. In discussing objectified noises we find for the first time the attempt to deal with circumstances that Katz had not discussed in any of his works; various stimuli (noise, smell, taste, and pain) can be combined to yield "thing-characteristics," but only if *intermodal* qualities are involved. On the other hand, the dimensions of reality, e.g., of "noise-objects," can be reduced, so that finally only the pure noise remains. As is commonly known, such "pure sensation" is never found in natural surroundings. It can only be experienced in laboratories as a result of artificial refinement. The objectivation of noise depends therefore either on optic or tactual stimuli simultaneous with the auditory quality. Nevertheless, sounds do have phenomenal quality. The tone-colors of different musical instruments are evidently not alike. The strongest loudness-invariance manifests itself in those tone qualities that are functionally connected with the most complex relation-systems, as in the case of the human voice (14).

Katz's fruitful idea can also be applied to smell, taste, and other types of perceptions. I have also distinguished phenomenal modes of appearance for these modalities. Henning (3) describes several objectivizing and resolving processes in the sphere of smell sensation. I have postulated *floating smell* and *objectified smell* as phenomenal modes, and correspondingly *swelling taste* and *objectified taste impressions*. A floating smell and a swelling taste can only become objectified with parallel optic and haptic stimuli, respectively. The crux of the question is found with warmth, cold, and pain perceptions. Warmth and cold may remain isolated in floating or condensed states, or can be objectified with simultaneous optic or haptic impressions. Pricking pain (but not other forms of pain) is objectified (1), in part, in terms of stimulation provided by stabbing instruments (needle, knife, etc.) because of the added tactile and perhaps also the form component.

Our phenomenal object-configurations are, therefore, for the most part intermodally determined. It is impossible for us to imagine a nonintermodal world of things. The different modal systems do not seem so hermetically sealed from each other as was thought in the days of J. Müller. The senses constitute a single system. For man, undoubtedly, the world of sight is the most important, but surely a purely visual world would be an abstraction. Such a thing is never found in empirical reality. It is, for example, impossible on the basis of visual impressions, when the skin is anaesthetized, to decide when and at what point the surface of the body becomes excited above the threshold. However, the perception of one's bodily orientation in a dimly lit room is a function of a visual ideational structure.

But there are also cases where the haptic and vestibular controls of the optical organization are in conflict. Only harmony between the input systems in question can restore the equilibrium of the perceptual field. Evidently these optic and haptic components possess special integrative functions for the "real" gestalts. The reality-impression of objects, their material and substantial character, is probably grounded in the repeated anchoring of phenomenal objects. These "moorings" arise from the coordination of the perceptions with basic optic and haptic elements, that is to say, from the process of objectivizing. "Rigid objects" seem to depend on cooperation of the visual and auditory senses. It seems evident that no gestalt occurs on the basis of taste, smell, vibration, warmth, cold, or pain impressions. Acoustic impressions can pass over into gestalten, but the time factor is then an organizing element. Temporal organization yields a dynamic, successive gestalt. In this, the "corticalized level" of the auditory sense plays a decisive role. Smell and taste do not have "melodies" because of the impossibility of object-gestalten at this level.

From the foregoing we can conclude that organized perception (that is to say, at the cortical stage) is really a system of relationships. Evidently we react to certain *relations* in which we—phenomenally speaking—experience contents. In this concrete relational environment there exist, to be sure, various specific stimuli. They exert their influence on our afferent nerves. That is the *physiological* side of the perceptual process. The psychological side of the process implies that the organism reacts as a whole to a relational system of stimulus-effects. These relational systems are almost unsurveyable in their complexity, e.g., with respect to the functional foundations of a perceptual object. Until recently we have been able to analyze quantitatively these complex wholes only with respect to some details of their appearance. Every stimulus is, to paraphrase Metzger, a stimulus-system, besides having its specific function. It also merges with the whole stimulus-system in its totality (11). From this fact it follows that we get firmer and more exact perceptual achievements, the more the sporadic stimulus-components concurrently connect themselves to a framework. In this sense the well known facts of phenomenal object-constancy (size, form, and color constancy) depend on a richly organized pattern of stimulus-relationships.

We have already mentioned Katz's statement about "field-size." The object-color seems the more "retinal," i.e., more nearly pure film color, the smaller the field in which it is presented. Katz could thus establish that the color constancy of perceived objects under several colored illuminations increases as the field of view either is enlarged or seems to be enlarged. Metzger has shown convincingly that these findings *do not* depend on the limitation of the *field of vision*. The opportunity to freely look around is, of course, important, since better possibilities of separate analyses of the constant chromatic components of the illumination are then at hand. Free observation with moving eyes affords the perception of numerous "systematic stimuli." But the impoverishment of this clue can also be obtained by kinds of reduction other than by cutting down the field of vision. Metzger's beautiful experiments with the homogeneous total field show that we can give to the field of vision *any* size without obtaining thing-constancy (10). The articulation of the stimulus-field is decisive. Minimal articulation can also be a consequence of too weak an illumination. In this way "reduction" of color can be brought about.

When we say that Katz is right, we must comprehend generally the validity of what he has formulated. It is valid for more than vision only. The concept of "reduction" is applicable to all senses for which objective patterning exists. The effect of the visual "reduction-screen" is really only a special case. We meet everywhere in all stimulus situations invariant perceptual objects, previously encountered and formed. And, analogously, we always find a declining degree of invariance when the reciprocity of the concomitant stimuli is reduced by any method. We must not think that this reduction can be obtained by identical technical means in all the different sense-spheres. The multiple stimuli for vision are obtained by the scope of the field of view or by the illumination of the field of view. For haptics, it is through the kinesthetic sense, for instance, the finger tips. For the invariance experiments in olfactory and gustatory perception, a reduction can only be obtained by reducing the possibilities of *objectivation* (through gradual elimination of vision and touch). The meaning of "articulation stages" can be very beautifully clarified by experiments on loudness-constancy with changing distance. The human voice is the most constant and least variable sound, while a buzzer-sound has the lowest invariance level (14).

We now leave this train of thought and occupy ourselves with some related points of view. To be perceived, objects must produce spatial location. The reaction to the "relational system" that has been indicated above predicts that we move into our perceptual products in certain invariant coordinations. The development of the subjective coordinate system belongs among the most interesting chapters of the psychology of perception. The whole field of perception possesses the same validities as we have established through the "unique thing." A "bare object" represents a peculiar if not limited relational system. The localization of the perceptual objects occurs in connection with the three orthogonally constructed interacting planes, the vertical, horizontal, and sagittal. The medium line of the perceptual frames is assigned to these coordinations with amazing accuracy. The phenomenal local signs, "right," "left," "before," "behind," etc., apply to this system. The picture which we experience to our right remains "the same" when we turn our eyes and locate it to "the left." By so doing, we have joined the time dimension to the coordinate system. The example of the picture demonstrates the invariance of the object in time.

Once more: Because the precise and clearly experienced external world supposes a mediating relational system in organized detail, its concreteness and stability are completely lost if we experimentally reduce the optical-acoustical-haptical stimulus pattern to a minimum. In experiments with movement perception (13) such loss can be demonstrated. If the subject, for instance, is told to estimate the speed of an evenly illuminated paper square which moves sideways with uniform speed in the frontal plane, one thereby finds considerable variance between the successive estimations when the experiment is carried out in a completely dark and "reference-poor" environment for the motion. If more richly organized reference-criteria are admitted (normal daylight, "passing things," structured background, and the like), one immediately judges with greater accuracy and with decrease in the variance of the observations. The largest overestimations of the speed are elicited by the most reduced configurations. One lone clue, a stable light source in a dark room, is often enough to bring about a considerable decline in the errors (about 15%) and the variance (about 5%).

Through another experiment at the Psychological Institute of the University of Helsinki, it was found that the subject succeeded essentially in a "localization task" with sound sources in a dark room if he could use a "reference sound source" which was attached to a slightly movable long pole between the subject and the "goal." The subject himself could "regulate" this sound-source (buzzer). Most effective was the "acoustical-optical method," whereby the pointer was also illuminated (14).

Of all the phenomena which produce a permanence in the frame of the space-time-coordinate system, our empirical self is probably the most important. We notice most distinctly the temporary interference and functional crises of the percepeual-psychological system in the not-too-rare cases where we wake up in unknown places, e.g., in darkness. All conceivable clues for the perceptual function seem to be absent—even the possible pressure experiences and kinesthetic impressions which are recalled by our position in bed. And nevertheless self-permanence, which we doubt least, shapes the perceptual content—apart from some pathological conditions. According to Gadelius (2), the empirical self will, through this temporal invariance, be constituted as a perception phenomenon with the perceptual components of the body-image. It is one of the highest acquisitions of perception. Psychology has doubtless shown that the invariance of the self-experience depends on conditions similar to the conditions responsible for the perception of *objects*. Kaila observes in this connection:

> The so-called "phantom limbs" of the amputated part of the body show what tenacious constancy and hallucinatory intensity this body-scheme has: a war invalid for instance, the one who has lost his leg, is initially exposed to danger because he experiences hallucinations of his lost leg, so that when he gets up out of bed, for example, he makes a step with the non-existing leg (5, p. 164).

With the most trivial and commonplace perceptual systems—the phenomenal objects—as a starting point, we have proceeded to the most complex thing that a human being is able to perceive at all: his empirical self. The main conclusion to be drawn from our discussion is obvious: It is the stimulus pattern as a system of relations, not as representing some "substantial stuff," which immediately gives rise to the impression of a three-dimensional world.

In summary, I believe the position represented here paves the way for modern perceptual psychology to arrive at functional principles of explanation.

REFERENCES

1. Buytendijk, F. J. *Über den Schmerz.* Bern: Hans Huber, 1948.
2. Gadelius, B. *Human mentality in the light of psychiatric experience. An outline of general psychiatry.* Copenhagen: Levin & Munksgaard, 1933.
3. Henning, H. *Der Geruch.* Leipzig: Barth, 1924.
4. Kaila, E. Logik und Psychophysik. Ein Beitrag zur theoretischen Psychologie. *Theoria,* 1944, **10**, 91–119.
5. Kaila, E. Emotion, vilja, personlighet. In D. Katz (Ed.), *Handbok i psykologi.* Stockholm: Bonniers, 1950.
6. Katz, D. Der Aufbau der Tastwelt. *Z. Psychol.,* 1925, Ergbd. 11, 1–270.
7. Katz, D. Der Aufbau der Farbwelt. *Z. Psychol.,* 1930, Ergbd. 7, 1–484.
8. Katz, D. *The world of colour.* London: Kegan Paul, Trench, Trubner, 1935.
9. Katz, D. *Gestalt psychology.* New York: Ronald Press, 1950.
10. Metzger, W. Zur Phänomenologie des homogenen Ganzfelds. *Psychol. Forsch.,* 1929, **13**, 6–29.
11. Metzger, W. *Psychologie. Die Entwicklung ihrer Grundannahmen seit der Einführung des Experiments.* Heidelberg: D. Steinkopf, 1954.
12. Révész, G. *Die Formenwelt des Tastsinnes. Grundlegung einer Haptik, und der Blindenpsychologie I-II.* Haag: Martinus Nijhoff, 1938.
13. Takala, M. Asymmetries of the visual space. *Ann. Acad. Sci. Fenn.,* **B72.2,** 1951.
14. von Fieandt, K. *Varseblivningens psykologi.* Stockholm: Natur & Kultur, 1956.

(Received April 10, 1958)

VOL. 65, NO. 6 NOVEMBER, 1958

THE PSYCHOLOGICAL REVIEW

HERBERT SIDNEY LANGFELD

1879–1958

In 1904 Herbert Sidney Langfeld resigned his position as Secretary to the Naval Attaché of the American Embassy in Berlin, for he had decided against a career in the foreign service in favor of graduate study in psychology at the University of Berlin. This decision brought a real loss to American foreign service, for in the very best sense Langfeld was a diplomat to the manner born; but at the same time it was an important gain for American education in general and psychology in particular. Langfeld was the last American to study under the first generation of German psychologists, and revealed during his long career in this country, both as teacher and administrator, a serene wisdom and detachment together with a breadth of outlook which may have resulted in part from his experiences in the German universities. The psychologists with whom he studied in Germany were also philosophers with imposing ideas and a sweep of imagination which often made up for their lack of knowledge regarding the experimental techniques and statistical analyses in which our students today are so carefully trained.

Dr. Langfeld received his early education at the Central High School of his native Philadelphia and at Haverford College. After he enrolled in the University of Berlin he did most of his work under Carl Stumpf and received his Ph.D. under him in 1909. He published several articles in the *Zeitschrift für Psychologie* in the field of visual phenomena, rather than in Stumpf's favorite field of audition. At the same time he showed a profound interest in art which remained with him throughout his life, and spent many hours in museums and in the classrooms of art historians and aestheticians. On returning to this country he accepted a position as instructor in psychology in Harvard College. When he left Emerson Hall fifteen years later he had served Harvard as associate professor and director of the laboratory for a number of years, had written articles on a variety of topics both within and outside the field of experimental psychology, and in 1920 had published his most distinguished and best-known book, *The Aesthetic Attitude.* He had also written, in collaboration with Floyd Allport, a laboratory manual which was widely used in various universities for a number of years.

Langfeld was a good lecturer, skilled in clear exposition and in the use of language "understanded of the people." He was not dramatic, nor did he ever lose his natural dignity by trying to be popular or obvious. The duller students were not impressed, but he had the teacher's reward of knowing that those who took the pains to follow and grasp his ideas retained an enduring and favorable impression. Tucked away in his desk Mrs. Langfeld recently found a petition signed by a large number of Harvard graduate students begging him to offer a special course in systems of psychology because of their conviction that he could give them a scope and depth of insight which they felt they badly needed. In a recent volume, *College in a Yard,* one of the many distinguished contributors who recalled their memories of famous teachers at Harvard wrote that "In psychology and aesthetics there was Herbert Sidney Langfeld, whose occasional ironies were as nice, as elegant and impeccable as his habit."

In 1924, on the completion of Eno Hall, Langfeld accepted a call to become Professor of Psychology at Princeton and to succeed Howard Crosby Warren as Director of the Psychological Laboratory. Warren had managed to separate psychology from philosophy only three years before, with the result that he and Langfeld were burdened with the heavy responsibility of building up a department faced with indifference, if not opposition, from the philosophers and the administration towards what William James once called "that nasty little subject." In 1932 Langfeld became Chairman of the Department and in 1937 was awarded the Stuart Professorship. During the years until his retirement in 1947 the department more than doubled in size of staff and in the scope of its activities. Langfeld became deeply devoted both to the town and gown of Princeton, and resented any criticism of the institution or the department almost as though it were a personal affront. On rare occasions he allowed himself the privilege of speaking with disfavor about minor matters relating to Princeton, but it was not good form for anyone else to do so in his presence. He was admired by graduate students and members of his staff, for they accepted his pride in their work as a challenge to their best efforts. They always referred to him affectionately as "Langie," but not when he was within earshot. When Wever and Bray made themselves famous by their studies in audition, he experienced the satisfaction and delight of one who was somehow sure that only in his laboratory could such important ideas be born.

Throughout his writings there appear numerous variations on the theme of a motor theory of consciousness, or, as he preferred to call it, a response psychology. This view has always been closely associated with his name, although in one of his articles he disclaimed any credit for formulating the theory and showed in impressive detail its origins and development from Greek philosophy down to the immediate predecessors of behaviorism. Undergraduates both at Harvard and Princeton were always warned by their tutors and preceptors that they should know the motor theory forward and backward, for no matter how varied the wording of an examination question, there could be only one correct answer. Langfeld was never able to believe, as did some of his German professors, that consciousness can be isolated for phenomenological study in its own right. It is too closely linked to efferent as well as to afferent events. The data of sense, as well as those of perception, feeling, cognition, and emotion have their terms interpenetrated and closely bound up with motor processes.

This view of mind, with its emphasis on motor response, led into and was then swamped by the behaviorism of the 1920s. With radical behaviorism, however, Langfeld had no sympathy; and it is gratifying to realize that he lived long enough to see his own view, albeit with new and more elaborate terminology, restored to its proper and appointed place in the psychology of today. He was often amused by the tortuous labor of those who had not fully recovered from the behavioristic revolt to find "objective" sentences for the description of what at one time would have been called simply direct experience or consciousness. He knew that consciousness by any other name would still be the same thing, and he continued to insist that the data of classical psychology are an inescapable part of the intricate chain of external and internal events that make up the totality of mental life, and that these events are best looked upon and labeled as neutral rather than either subjective or objective. His refusal to become involved in any sort of body-mind controversy was undoubtedly reinforced by his close association with the philosopher-psychologist, Edwin B. Holt, who frequently argued that attempts to tell what is objective and what is subjective are a waste of time for the simple reason that they never get anywhere.

In his writings on aesthetics, and especially in *The Aesthetic Attitude,* Langfeld drew heavily on the doctrine of empathy for an explanation of the very special and affective perceptual qualities of art. Hardly anywhere in English can a more persuasive account be found of the manner in which incipient muscular movement may conceivably be the basis of the judgment that a work of art is exciting, wistful, melancholy, sprightly, or what not. The problem is among the oldest in speculation about art and has given no end of trouble to philosophers, art critics, and psychologists. It seemed absurd to argue that the perceived object of art itself could be melancholy; and there was good evidence to show that those who described a work of art in such manner were not ordinarily in a melancholy state. Here was a problem, for if neither the object nor the subject is melancholy, where does the quality come from? Langfeld seized upon Lipp's doctrine of *Einfühlung* and worked it for all it was worth. Aesthetic qualities must somehow have their origin in the subject, but are immediately and unconsciously projected upon and identified with the object, so that it is the object which *seems* to be melancholy. The notion is an obvious variation on the theme of a motor theory of perception, and was therefore doubly congenial to Langfeld because of its close relation to his more general point of view in psychology.

In recent years, however, the theory of empathy has been severely attacked, and Langfeld was placed in a much more defensive position with respect to many of the arguments to which he was deeply attached. The implications of *Gestalttheorie* for art, especially in some of the devastating paragraphs of Koffka and Köhler, were drawn upon by a number of philosophers and psychologists to reveal the serious limitations of empathy. The doctrine gives too little credit to the genius of the artist. It implies that the subject creates from within himself the glories of the glass of Chartres, the majestic power of the sculpture of Michelangelo, and the wistful melancholy of some of the minuets of Mozart. The artist merely puts together a sort of neutral perceptual design and then waits for the subject to perform the act of artistic creation. Because this act seems too much to hope for, the pendulum has swung in the direction of assigning to the artist the ability to create configurations which themselves possess the *tertiary* qualities of longing, serenity, erotic passion, sadness, gaiety, etc. Sadness may be present in visual and auditory as well as in organic impressions. It is this possibility of multiple location that has caused the confusion in aesthetic theory. A person may feel sad. A piece of music may sound sad. But neither state necessarily depends on the other. It is not empathy which gives to the slow movement of the *Eroica* the sense of poignant grief. It was Beethoven who performed that miracle, and for those with ears to hear, it is a source of pleasure rather than of grief, although the judgment may be made that the music is profoundly sad. Yet the doctrine of empathy is still widely held, and may in time recover from its wounds. There are few views in philosophy and psychology which are completely down and out. In any case, Langfeld's treatment of *Einfühlung* is masterly and for a long time will remain required reading for every student of aesthetics, for even if a student entertains doubts about empathy he must know how to answer the penetrating arguments which Langfeld has marshaled in its support.

Langfeld's eclectic interests in psychology and his respect for good writing made him an ideal editor. As the years went on, he spent an ever increasing amount of time in going over manuscripts with painstaking care for professional journals and for the publishing house of John Wiley and Sons. In 1931 he was appointed editor of the *Psychological Monographs,* and then for thirteen years served the *Psychological Review* in a similar capacity. He was joint author and editor of a number of texts, the best known of which was the famous *BLW*, a factual foundation for psychology which he, Boring, and Weld prepared for such advanced introductory courses that many graduate students have used it to bone up for their final oral examinations. His services to psychology have been recognized in a variety of ways. He was an honorary member of the Belgian and Swedish Psychological Associations, and in 1930 was elected president of the American Psychological Association. His election to the Pontifical Academy of Sciences in 1949 and the private audience with His Holiness Pope Pius XII gave him, an Episcopalian, deep pleasure and satisfaction, as did also the award of *Docteur de l'Université* which he, Michotte, and Piéron received from the University of Montreal at the time of the 14th International Congress of Psychology.

Throughout his long professional career Langfeld provided a valuable link between psychology in America and psychology in other countries. He was Chairman of the Committee on International Cooperation in Psychology of the National Research Council, and Permanent Secretary of the International Congress of Psychology, and in his many travels abroad he established close personal and professional relations with investigators in other countries. It is safe to say that no American maintained such friendly contacts with so many psychologists in other lands. Whenever his professional friends had a chance to come to this country, they hastened to accept Langfeld's invitation to visit him in Eno Hall and to partake of the hospitality so generously offered by Dr. and Mrs. Langfeld in their Princeton home.

Retirement made little change in the routine of Langfeld's life except for the freedom from classroom instruction and administrative responsibilities. He came regularly to his office in Eno Hall, carried on voluminous correspondence regarding professional matters all over the world, wrote articles and prepared lectures to be delivered before learned societies in this country and abroad, watched with deep interest but without any interference the activities of the department over which he presided for so many years, and continued to extend the hospitality of his home to graduate students, members of the staff, and especially to visitors from abroad. Shortly before his death he spoke informally at one of the regular seminars in Eno Hall. He left an indelible impression both with staff and students of his ability to expound clearly the broad historical background for his sharp comments about contemporary psychology.

CARROLL C. PRATT
Princeton University

Psychological Review
Vol. 65, No. 6, 1958

PSYCHOLOGICAL STRUCTURE AND PSYCHOLOGICAL ACTIVITY[1]

HELEN PEAK

University of Michigan

The science of behavior, like all science, is preoccupied with relations between events. For this reason, it is important to analyze the diverse relational concepts that have been employed in describing psychological structures, in order to discover whether unique characteristics belong to concepts such as similarity, opposition, complements, antecedents, and consequents, or whether they are all reducible to simpler and more fundamental relational properties. Another paper (**28**) has suggested that all kinds of psychological structures can be described in terms of parts or events standing in spatial and temporal relations to each other, the space being conceived in psychological terms; and that every arrangement implies the course that activity will take when the structure is activated at specified points. It will be the task of the present paper to elaborate this analysis in somewhat more detail and to suggest how selected structural patterns influence the course of events in perceiving and acting.

The properties of structure will be described and certain basic principles about the relation of structure and activation will be proposed. These will be discussed as they apply to phenomena such as context effects, assimilation and contrast, the algebraic summation of positive and negative attitudes to produce intermediate attitudes, judgment within "balanced" and "unbalanced" structures as described by Heider (**11**), and the persisting activity of the motivated state.

PSYCHOLOGICAL STRUCTURE

If the course of psychological events is to be predicted from information about the activation pattern imposed on a known structure, the meaning of the central concepts involved must be made clear, most particularly the notions of structure and activation.

Stated briefly, activation is a term similar in meaning to stimulation but more general in the sense that the change which *activates* a structure may be either the energy change which takes effect by way of sense organs (a stimulus) or a central event, such as an aroused concept, which in turn produces change in or activation of another structure. Activation is transmitted change.

The term *structure* is used in a very broad sense to refer to a system of relationships between identifiable parts. It is a hypothetical construct[2] to be inferred from events observed under specified conditions and, like all such constructs, once inferred, it becomes the basis for predicting behavior on subsequent occasions when there are reasons to believe that the system of relations has not changed. For example, if it is ascertained that under free association instructions the stimulus word *black* elicits the response *white* 80 percent of the time in some population or in some individual on a series of occasions, then it is inferred that there is a relationship between the stimulus-aroused concept and another part of the structure whose activation produces the observed response. This familiar fact that organisms retain dispositions to react in certain ways must imply that the underlying structural arrangements have a basis in the nervous system. However, it is not our purpose to present a neurological model. Rather the intent is to describe certain patterns of relations which are compatible with important psychological phenomena. The neurological substrate of these phenomena may or may not possess characteristics of the kind to be described.

---

[1] This article was written in connection with ONR Contract Nonr-1224(10) NR 171-039 with the University of Michigan.

[2] The term has a less restricted meaning in this paper than is given it by Meehl and McCorquodale (22). The ideal is, of course, to define any construct rigorously both at the conceptual and at the operational level in order that the same structure will be inferred from the various events to which it is assumed to contribute. But the degree of rigor of the definition does not determine whether it is a construct or its hypothetical character.

Now the components of a structure will vary not only with the subject matter of the particular science but also with the theoretical preferences of the scientist. But parts of some kind must be identified, whether they be called stimuli, roles, concepts, regions, R_g, attitudes, cathected objects, or drives—to mention a few miscellaneous constructs that have been regarded as segments of the matrix of events by equally miscellaneous psychologists. Moreover, it should be noted that the components of a structure are not to be regarded as indivisible elements, for some components may be highly structured with relations existing between the parts, as in the complex concept of a friend, while others are relatively simple and homogeneous, as would be the concept of warmth. This means that there may be relations *within* components parts as well as *between* parts of a larger structure.

It has already been noted that the nature of the relations existing between the parts of a structure will be a principal concern of this discussion and that the effort will be made to conceptualize the more complex relations in terms of the concepts of position and distance. Therefore, these basic relational notions must now be defined.

Position

To assign position means, generally speaking, to place an event or a point in relation to one or more other events. In the simplest case, position depends on the process of identifying and distinguishing an event from other events, and to have the same position means simply to be in the same category.

But position may be specified more precisely if events have been ordered on a continuum or arranged in a series which derives its order either from the amount of a property or from the temporal order of activation of the individual parts of the series when one part is activated. In other words, position may be stated in terms of membership in a category (nominal position) or in terms of some amount of a property or position in a series (ordinal position, or a point on an interval or ratio scale).

Distance

The psychological distance from any point A to point B is defined in terms of the number of units or steps in an ordered series of some kind which intervene between A and B which are themselves part of the series.

If, then, two points have only nominal position in separate categories, their distance apart will be indeterminate unless the categories themselves can somehow be ordered.

In other words, the description of a distance must involve finding a method of ordering events before it can be established (*a*) that other points of a series do in fact lie between *A* and *B*, and (*b*) that there is some determinate number of such intervening events that define the distance.

Further, it must be noted that psychological distance has a special property which makes it different in this respect from physical and mathematical distance: it need not be symmetrical. The distance from *A* to *B* is not necessarily the same as from *B* to *A*. For example, the occurrence of an event which is an antecedent will more often elicit the consequent as an associate than presentation of the consequent will elicit the antecedent. It is inferred that the distance is greater from consequent to antecedent than in the other direction.

Although this is not the appropriate place for a discussion of all the operations for determining order and distance among psychological events, a few examples of operations from which psychological distance has been inferred will make it clear that all imply the foregoing definition. Distance between two points may be inferred from judgments of the position of each point on a scale. The subject matches his reaction to the stimulus objects to some scale position and the difference in position represents the number of intervening units. Or distance may be judged directly where the subject is asked to state how far apart two stimuli are. For example, he judges how different John and Jim are in intelligence, and reports it. The units used in the judgments and their distance apart may or may not have been scaled by techniques such as those developed by Thurstone (36), Coombs (8), Guttman (35) and others, but all these methods imply the occurrence and identifiability of intervening steps of some kind. Furthermore, the relative frequency and the latency with which one event follows another have also been considered to reflect distance between events. Indeed, probability of activation of one event by another is often taken as synonymous with distance (26).

If it be assumed that each of these measures reflects in some degree a distance between A' and A'' in psychological space, the major operational problem is to find ways of removing the effects on behavior of determinants other than distance A'–A'', for it is only when these irrelevant determinants have been partialled out of each operation that the various operations for inferring this distance can be made comparable.[3]

Direction

Direction involves determining position of some event with respect to two or more other positions.

The course of activation can be said to have direction, but there is a question whether direction is a necessary concept for the description of structure. Suppose that I know three boys, Tom, John, and Jim, whose IQs are 100, 110, and 115. It is possible to say that John's IQ lies in the direction of Jim's from Tom's, and that Jim's and Tom's are in opposite directions from John's, in my psychological structure. But this description in terms of direction adds little to a statement in terms of position and distance. It is when predicting the course of activity within a structure that the concept of direction becomes useful, for it is necessary that information about the points of input into a structure be provided, in addition to knowledge of the structure itself, before the direction of events can be determined. It is the relation of subsequent events to the initial point of activation or the input point that gives direction.

---

[3] See Hilgard and Marquis (12, p. 135) on the low correlation between indicants of habit strength, such as the probability of activation of *A* by *B*, judged similarity of *A* and *B*, generalization of response from *A* to *B*, speed of extinction, latency, etc.

ACTIVATION

To say that a structure has been activated means that change from some source has set off change in a structure. Thus the energy change that is a stimulus event, such as the spoken word *good*, initiates a successive wave of change in parts of a complex structure finite distances apart until the process leads finally by a series of activations to the spoken word *bad*. Potentiality for this series of events implicit in the structural arrangement has been actualized by the transmittal of change from one point in the structure to other points.

In some cases activation produces only a temporary disruption of the structure which tends to be reconstituted following activation. Thus a violin string is distorted when a bow is drawn across it but soon returns to its original state. In a somewhat different way, neurone structure is partially destroyed in responding to stimulation but is polarized again after the passage of an impulse. Again, the activation of a series of concepts, percepts, movements, or word responses involves structural change, but here, too, activity subsides and the original state of quiescence returns.[4]

But activation may leave an altered structure, which means that the structure does not invariably return to the preactivation state. This is known to be the case when later activation at the same point, under conditions otherwise the same, produces a different outcome. Restructuring or learning has occurred. The presentation of a complete learning theory is beyond the scope of the present discussion, which addresses itself primarily to the problems of the immediate effects on perceiving and acting of activating a given structure at a given point or points. However, it may be said that any condition which has the effect of activating two structures close together in time increases the probability of a change in the relationship between the structures. Furthermore, the shorter the existing distance between the structures when activated, the more likely it is that their simultaneous activation will reduce the distance still further. The role of affective consequences in learning cannot be considered at this time.

ACTIVATION PRINCIPLES

Structure is of interest only insofar as it determines the course of activity. The next step is, therefore, to analyze certain assumptions that must be made about the relation of activation and structure. Five principles are proposed that appear to be necessary in order to derive theorems about activation patterns that develop within particular structures. The validation of these constructs will depend upon the success with which they generate propositions which are verifiable. At present they derive their justification from the fact that they appear to be compatible with certain phenomena for which there is observational support.

Distance and Activation

It is assumed that the probability of activation of some point *B* when *A* is activated is an inverse monotonic function of the psychological distance between A and B, when other sources of the activation of B are excluded (Principle I).

---

[4] The disruptive events of activation and the opposed tendency to reconstitute structure suggest entropy principles, activation corresponding to positive entropy and reconstitution of structure to negative entropy (**33**).

As already explained, the activation of any point in psychological space may be determined by input from more than one source, and it is only when activation of B from other sources is controlled that this relation can be assumed to exist between distance A–B and the probability of activation of B by A.

As a matter of fact, it must be the exceptional occasion when input to any point in a complex structure comes from only one source, and so it is necessary to make assumptions about the way in which activation from different sources combines to determine the next steps in the activation pattern. There appear to be two basic patterns which demand different combination principles. In the first, a serial arrangement, activation of some structure at the end of the series depends on the occurrence of activation in preceding steps of the series. In the second pattern, structures are arranged so that there is multiple input to the same point from two or more sources.

Serial Arrangement

In a serially arranged structure, the probability of activation of some point C at the end of the series when a point S is activated is the product of the probabilities of activating points intervening between S and C (Principle II).

For example, in Fig. 1 the probability that S_a will activate C equals 0.63 (0.9×0.7). The probability that S_b will activate C equals 0.36 (0.9×0.4).

Multiple Input to the Same Point

The probability that any point will be activated is a function of the combined probabilities of its activation from all currently active structures. Given the probability that A will activate C and that B will activate C, the probability that A and B together will activate C is assumed to be equal to:

$$1 - (P_{CA} \times P_{C\bar{B}})$$

where P_{CA} is the probability that A alone will *not* activate C and $P_{C\bar{B}}$ is the probability that B alone will *not* activate C (Principle III).

For the values given in Fig. 1, the probability that A and B together will activate C is equal to 0.77; that is, the combined probability is greater than the probability produced by either component (A or B) except when the probability of one or both terms is zero or unity.

The logic of the assumption is simply this. The probability that C will occur when both A and B are activated is the sum of those occasions where A activates C but B does not $(P_A \times P_{\bar{B}})$, where B alone activates C but A does not $(P_{\bar{A}} \times P_B)$ and where both activate C $(P_A \times P_B)$. In other words, the relative frequency is reduced by the frequency of those occasions on which neither A nor B activates C (probability that C not activated $= P_{C\bar{A}} \times P_{CB}$). This is the term subtracted from unity in the equation.

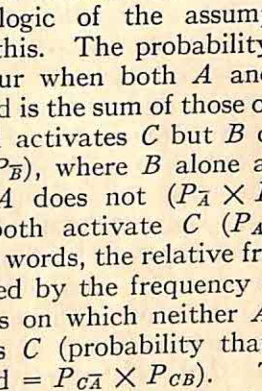


FIG. 1. Serial arrangement of parts: $S_a \rightarrow A$, $A \rightarrow C$; $S_b \rightarrow B$, $B \rightarrow C$. Multiple input to the same point: $A \rightarrow C$ and $B \rightarrow C$.

Examples of situations in which complex multiple input may be assumed to occur are numerous. The many studies which emphasize the effect of preparatory events, such as context, set, instructions, and perceptual response dispositions on reactions to stimuli, belong here. Howes and Osgood (**15**) showed, for example, that both the number of preparatory stimuli similar to a word and the interval by which those preparatory stimuli preceded a stimulus word influenced the nature of the free association response made to the word. The preparatory stimuli may be assumed to have activated the same structure as did the stimulus word, and thereby to have increased the probability of the occurrence of a response depending on activity in that structure. And Asch's well-known demonstration (**1**) of the effects of prior presentation of trait terms, such as *warm* and *cold* on the interpretation of other traits attributed to an individual, may illustrate the same point.

It appears that the influence of central motivational states on perception and action may likewise be conceived in terms of multiple input. Many examples come to mind, such as the familiar studies of Atkinson and McClelland (**2**), in which TAT stories were written under different degrees of hunger motivation, with the consequence that references to eating and/or activities instrumental to getting food increased with increased time of deprivation.

Again, the apparent increase in the breadth of the generalization gradient when drive level is high or with high anxiety may also involve this same principle, if it is assumed that anxiety or high levels of motivation set up diffuse activation and hence multiple input to all points on a gradient. Finally, it should be noted that the process of abstraction should also be examined with this principle in mind. A property common to different situations is more likely to be noticed if an individual experiences a number of different events all of which possess the property. Thus the European who sees ten Americans chewing gum is more likely to abstract a concept of gum-chewing Americans than if he has only one such experience. The multiple input to the gum-chewing property increases its probability of arousal.

Event Units

Bridgman has aptly described the problem of setting boundaries to physical events when he says, "I believe the essence [of the identification of an event] lies in the discrete character of the event. . . . An event is by definition the aggregate of all phenomena bounded by certain discontinuities." He then goes on to point out that this concept can only be an approximate one, because there are no such things in experience as sharp discontinuities. Rather, as measures become more refined "the edges of the supposed discontinuities become blurred" (**5**, pp. 95–96).

This is the problem of discovering cleavages in the matrix of events, the boundaries which result in some events being part of the same whole while others are parts of different wholes. This search for gaps between phenomena is involved in delimiting the boundaries of perceived objects or of concepts, in determining what is included in a trait complex and what is excluded, in bringing it about that events will be "assimilated" or "contrasted."

It is proposed that the discontinuities that establish boundaries in psychological space are of two kinds: (*a*) the distance between events, decreasing distance tending to increase the likelihood that events will be part of the same whole or within the same boundaries; and (*b*) the probability of simultaneous arousal, the likelihood of events being included in the same whole increasing with the probability of their simultaneous arousal (Principle IV).

This means that anything which increases the probability of activating either or both of two points will increase the probability of their occurring together, for the higher the probability of their occurrence, the more likely they are to occur together.

Duration of Activity and Action Decrement

The repeated activation of the same structure at short intervals leads to activity decrement (Principle V).

The spread of activation will depend not only on distances between the parts of a structure and the pattern which determines the occurrence of multiple input at certain points, but it will also be influenced by the state of the to-be-activated structure. There is considerable evidence to indicate that the repeated activation of the same structure tends to produce a temporary reduction in activity; that is, activity results in some sort of change in the state of the system so that it is less reactive. This is what Lewin called *satiation* (**21**) and Hull, *reactive inhibition* (**16**). Köhler has provided evidence of figural aftereffects which points to a similar process (**19**). And more recently Walker and his associates have also confirmed the existence of a decrement in activity as a result of activity (**37**).

It would appear in general that the same conditions which result in developing clear foci of activity and sharp boundaries (IV) will tend to speed the onset of decrement. This suggests that short distances between the parts of a structure should lead to decrement, provided either that the parts are interrelated in such a way that they activate each other, or that input to the parts from some external source continues so that parts are activated repeatedly and simultaneously. However, the persistence of activity within a structure is not simply determined, for two consequences of distance between parts work in opposed directions. On the one hand, repeated activity at the same point is more likely to occur at short distances between parts, and hence to produce decrement. But it is also true, on the other hand, that great distances may reduce the probability of activation so much that activation ceases. The condition of maximum duration of activity is to be expected, therefore, at some distance between parts which is great enough to prevent immediate onset of decrement and yet small enough to provide relatively high probability of continuing transmission of activation from one part of the structure to another, under the existing conditions of input to the structure.

The application of this principle suggests that a simple stimulus object like the conventionalized drawing of a face would yield not only a simpler and more circumscribed activation pattern than will the actual presence of a person, but that the former stimulus should result in a more prompt perceptual act and a more rapid onset of decrement, so that the perceptual act should drop from "attention" more promptly.

THE DESCRIPTION OF SELECTED STRUCTURES

We turn now to the task of describing selected structures that are implied by familiar relational concepts such as similarity, opposites, complements, antecedents, and consequents. These classes of relationship are not sharply delimited nor have they always been defined in the manner here proposed. We believe, however, that the structures to be described are common ones, and that the particular terms used as labels are not without precedent. As already indicated, it is our intention to describe these patterns in terms of the concepts of position and distance, as previously defined. And in each case, the question will be: How does this arrangement affect the course of activation when the structure is aroused at different points? Obviously, a very large number of structural arrangements and of input patterns may be conceived. It has been necessary, therefore, to limit the task to a description of the structures mentioned above and to a consideration of the course of activity in such structures when certain simple input patterns are operating: activation at one point; at two points different distances apart, resulting in such outcomes as assimilation, contrast, or algebraic summation of opposites; and at three points, which will produce resultants of the kind that Heider has called balanced and unbalanced (**11**).

Similarity Structures

It has been common practice to define similarity conceptually in terms of the sharing of common properties. Sometimes this means that similarity between two events is described in terms of their distance apart on the same continuum. Again, it may refer to the number of continua (or properties) that the events have in common (**6**). Obviously some combination of the number of shared properties and the amount of each such property involved in the complexes would provide a more sensitive index of similarity than either alone. However, we shall consider at this time the simple conception of similarity defined in terms of distance between points on a continuum, for the more complex structure raises other problems, such as the organization of the separate properties within the complex, which must be given more detailed elaboration elsewhere.

To say that a common property is shared by two events means that the events lie on the same continuum. And in the words of Poincaré, ". . . a system of elements will form a continuum if we can pass from any one of them to any other by a series of consecutive elements such that each is indistinguishable from the preceding" (**30**, p. 52). In such a series some position *A* is indistinguishable from *B* and *B* from *C*, although *A* is not equal to *C*.

It is assumed that distances are short within a continuum, so that the probability of activation at any point in the continuum by activity at other points tends to be high. However, as the continuum comes to include more distinguishable positions, and hence to imply a greater number of events between the activation of one end and the transmission of the activation to the other end, the distance will be increased.

Now it should be recognized that the property represented by a continuum may be anything at all which the individual has abstracted from his experiences and structured in this fashion. This means that it may be blueness or brightness or other properties assigned to physical objects, but it may also be any of the qualities that we attribute to human beings, such as friendliness, conservatism, honesty, coldness, or aggressiveness, any one of which may be regarded as a simple property varying in amount and made up of a series of ordered positions without gaps between them.[5]

Opposites

At some stage a continuum may become so highly differentiated that the distance between the end points is greatly increased and subcategories develop within the more inclusive continuum. When this has occurred, moving from one end of the continuum to the other involves passing through a zero point or boundary which separates the emerging subcategories. Thus warm and cold within the dimension of temperature, or clean and dirty within the continuum of cleanliness, exemplify emerging subcategories which are in contact at a zero point. Such subcategories are usually called opposites.

The emergence of the opposites relation from a similarity continuum is implied in studies which have revealed that this relation appears relatively late in the individual's life history, as compared to similarity (**39**, p. 344). It is also of interest in this connection that various writers, including Freud (**9**), have noted that the same word may serve to represent a property and its opposite in ancient languages, such as Egyptian or Latin, while modern languages tend to use different words for these relationships. Latin provides examples like the word *altus*, meaning both *high* and *deep*, or *sacer*, to be translated as *holy* or *accursed*. Primitive speech is said to refer rarely to opposites, though relations of similarity are frequent (**38**), which again suggests the derivative nature of opposites relations.

But it must be recognized that opposites structures may also develop when a common property is abstracted from two initially unrelated continua or categories which have become associated to form subcategories within the abstracted dimension. The affective continuum, made up of positive and negative affect, may possibly emerge in this way, although the facts are not clear. Neurologically, positive and negative affect appear to depend on separate centers (**24**), and this suggests, though it does not prove, that the relation between them has been constructed by learning rather than initially given. On the other hand, early emotional reactions have been said to be undifferentiated excitement with the pleasant (approach) and unpleasant (avoidance) patterns emerging at a later date (**4**, p. 87).

Complements

Whatever the origin of the pattern of opposition, it may be regarded as intermediate in its characteristics between the similarity structure, within which distances are shortest, and those structures called "complementary," where distances are greater. The latter are composed of parts which are not necessarily related by a continuous series but which are nevertheless separated by some finite distance. Moreover, together the parts constitute a whole or they produce some common consequence. This is to say that complements are associated, even though not similar in the sense of lying on the same continuum and sharing a common property. Thus a figure and the background without which there is no figure, the concepts of the role of leader and followers, of men and women, may be thought of as pairs of events whose underlying relationship is usually of this kind.

---

[5] It is irrelevant for present purposes that the structure of personality traits is in fact more complex and that simple continua are rare. People may actually have such oversimplified beliefs about personality structure even though the psychologist regards the belief as inaccurate.

Antecedents and Consequents

When one event is responded to as preceding another, the psychological distance between them is greater than that between complements. The magnitude of this distance is a function of the length of the temporal interval between the occurrence of the events in past experience, as well as of the frequency and consistency of this experienced relation. Moreover, the distance from antecedent to consequent will typically be shorter than that from consequent to antecedent. This assumption is based on the fact of the relatively greater strength of forward than of backward associations (**13**, p. 622).

There are many situations that presumably involve structures of this kind. For example, a man and the acts which he performs will sometimes take on the antecedent-consequent relation. Heider (**10**) has presented interesting speculations about this matter, and has suggested that although these events may be a sequence, the man and his acts tend to be treated as a whole: the acts are characteristics of the man rather than separate events. We would suggest that it is necessary in any particular case to determine the distance between the two events, the probability that the activation of the concept *man* will elicit the concept of the *act* in question, as well as the probability that the activated notion of his *act* will elicit the concept of the *man*. As the distance becomes greater and the relation more asymmetrical, it will take on the properties of the antecedent-consequent relation. It is evident that precise boundaries between the complementary relation and the antecedent-consequent relation cannot be established. In this, as in all distinctions between the structures described, we are dealing with relative distances, so that it is simply a matter of convenience that leads us to apply different terms to relations which involve different relative distances.

Abstracted Relations

The discussion has been focused so far on the nature of certain concrete relations between events, but it is important to note that such relations may be conceptualized. Boring is making this distinction between a relation and the concept of it when he remarks, "A succession of ideas is not equivalent to an idea of succession" (**3**, p. 575). The idea of succession is an abstraction of one aspect of an activation pattern in which some event follows another. In the same way, similarity between two events is not the same as the idea of similarity. A simple similarity structure means that two points share a common property, in the sense of lying on the same continuum. But the pattern resulting from activation of two such points on the same continuum may have become related to another structure which, when activated, is the conceptualization of similarity. An abstraction is, therefore, just another event, resulting from the activation of a structure, but it is important to remember that distances within a *concept* of similarity-between-A-and-B will not necessarily be the same as the distance between A and B on the dimension from which this abstraction was derived. This does not mean that principles governing a certain relation and the concept of that relation are different, but rather that the precise structure of a concept of a relation must be determined, and not assumed to be the same as that of the referent structure.

Complex Structures

In order to make as clear as possible the essential features of the concept of structure as here presented, similars, opposites, complements, antecedents, and consequents have been treated separately as though two given events were related in only one of these patterns at a time. The fact is that the different aspects of complex structures may have different and multiple relations. Thus, the concepts of leaders and followers, though complementary, may also share common properties, such as having similar attitudes toward certain policies. Moreover, two complexes may share more than one dimension or they may have more than one complementary property. This means that the component aspects of two complex events may be different distances apart, so that the distance between the complexes themselves must be a function of the combination of these multiple relationships. It is suggested that the commensurability of the distances on different dimensions depends on the combination of the probabilities of activation implied by each relationship. In other words, if complex structures, A and B, have positions X_a and X_b on continuum X and positions Y_a and Y_b on continuum Y, then the distance from A to B is a combination of the probabilities that activation of X_a will result in activation of X_b and that Y_a will produce activation of Y_b.

ACTIVATION WITHIN SPECIFIC STRUCTURES

Assimilation and Contrast

The activation principles which have been stated must now be applied in order to illustrate the kind of resultants that may be produced within specific structures when activated at certain points. The first notion to be discussed is the familiar one[6] that at short distances between the points of input the result will be assimilation, and at somewhat greater distances, contrast.[7]

Despite the different shades of meaning that have been given the term *assimilation*, all uses emphasize similarity between processes in one way or another. Sometimes assimilation is known by the fact that the classification of events is the same; again, it is regarded as the incorporation of one event by another (**29**, p. 6); or it refers to the fact that two stimuli produce the same response (**31**, p. 8); or events are perceived as the same (**20**, p. 95). We shall apply the term to that resultant of two-point activation which is intermediate in some respect between the results that appear when either one of the points is activated without the other. In other words, assimilation has occurred when distance between the points is reduced by their simultaneous activation, so that a tall man looks smaller and a small man taller if they are together than when alone. On the other hand, contrast implies that the distance is exaggerated when two points are activated at the same time, with the consequence that the tall man looks taller and the small man smaller when they are together.

The recent studies of Sherif, Taub, and Hovland provide one illustrative test of the hypothesis under conditions which systematically vary the distance between structures activated, in order to determine whether or not there is a shift from assimilation to contrast as the distance increases (34). A series of weights was judged, first without an anchor and later following an anchor weight. In one condition, the anchor was equal to the heaviest of the original weights, and in another it was equal to the lightest. The judged heaviness of the weights *increased* when the heavier anchor preceded the weight being judged and decreased when the lighter preceded. This is the result that would be expected if the perception of either of two events, with a small distance between them ,lies between the position of the two judged separately. In another portion of this same study the authors showed that as the difference between the anchor and the weights of the original series was made still greater by an increase in the weight of the anchor, the judged weight of the original stimuli *decreased*. When the anchor was farther away from the weights in the original series and lighter, the original weights were judged heavier. This is the common observation of contrast made in the context of adaptation level theory.

---

[6] See, for example, Woodworth (**39**, p. 507; 569), Koffka (**17**, p. 502), Krech and Crutchfield (**20**, p. 95).

[7] Under conditions that cannot be elaborated, summation may possibly occur at the shortest distances. In this case either stimulus would be judged greater in the presence of the other than when alone.

Further evidence has been presented by these authors suggesting that reactions to communications about social issues follow these same principles, for statements about prohibition which fell within short distances of the individual's own position were seen by him as closer to his views, while those beyond this range were seen as farther away from his own position "than was actually the case" (14). In other words, there was a disposition to exaggerate the distance between own view and that of others in the latter case (contrast) and to underestimate it in the former (assimilation).

Although this kind of evidence supports the notion that distance between activated points is a factor in determining whether assimilation or contrast occurs, it remains to show why this should be so in terms of the assumptions made in this paper about distance and activation. We must ask what differences in the activation pattern would make us predict assimilation at short distances and contrast when the distances are longer.

The theory. The essential feature of the theory being proposed to account for this relationship is first of all the fact that at short distances simultaneous activation of two points A' and A'' by stimuli S' and S'' will result in multiple input to, and therefore maximum probability of arousal, of points lying between those positions that would be activated by the presence of either S' or S'' alone. This effect of multiple input at short distances follows from Principle III. Such a pattern of activation should result in a judgment of A' or of A'' which lies between the positions that they would take if presented alone.

When the positions of A' and A'' in the ordered series are farther apart, the number of intermediate points with multiple input decreases so that the activation surface[8] or pattern of probabilities of activation becomes bimodal, with two points of maximum probability of activation separated by points with lower probabilities. In this case, the judged position of either A' or A'' will be a function not only of the points of maximum activation on the surface but also of the number of events judged to intervene between the two maxima. According to Principle IV, the number of events between any two points is an inverse function of the probability that adjacent points will be activated simultaneously and a direct function of the distance between the points, when that distance is inferred from the position of the two stimuli presented at different times, so that there is no multiple input. It will be shown that at longer distances where the resulting surface is bimodal but still determined in part by the multiple input to intermediate points, the number of events, and hence the distance, will be increased disproportionately. This exaggeration of the number of events intervening between points of maximum activation results in contrast because the positions of the two maximum points will be seen as farther apart or separated by a greater number of events.

---

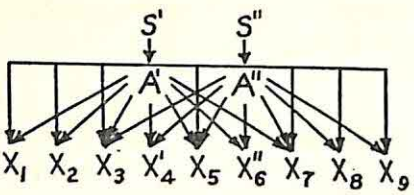
[8] From this point on, the phrase *activation surface* will be used to refer to a pattern of probabilities of activation resulting from input to a given structure at certain points. The line labelled *combined probabilities* in Fig. 2 and 3 represents the surface resulting from activation of the structure at points A' and A''. Similarly, the curves of Fig. 4 and 5 represent activation surfaces.

Development of activation surfaces as a function of distance between points of input. In order that this theory may be made clear it is necessary to show how different activation surfaces arise as a function of distances between the relative positions of input on some dimension. In Fig. 2, 3, 4, and 5, S' and S'' are stimuli which activate the continuum, A, at points A' and A'' respectively. Xs are other structures related to A at various distances. Now Xs, like As, represent any psychological process whatsoever—concepts, names, affective responses—that are activated with some probability by the arousal of A; they may or may not constitute a continuum; and they will be related to each other in varying degrees, depending on their developmental and activation history. However, if A is a continuum with the result that distances within A are short, and if distances between As and Xs are short, it should follow that Xs will have tended to become an ordered series connected by relatively short distances. This must be the case because Xs will have frequently been aroused at the same time as a result of activity in A. As and Xs in the accompanying figures are, then, simply different parts of the structure which are represented as having asymmetrical relations (connections in the direction from A to X). This means that activation of A will be followed by activation of Xs.

Figure 2 presents a schematic picture of relations existing between S' and A', between S'' and A'', and between points on A and certain Ss. It will be noticed that a given A point is represented as being related to several Xs. X'_4 and X''_6 represent direct, maximum probability responses to the activation of A' and A'', respectively. The connections of A' to X_1, X_2, X_3, X_5, X_6, and X_7 and of A'' to X_3, X_4, X_5, X_7, X_8, and X_9 are structural relations whose activation produces generalized reactions that have been developed in the following manner. Since distances within a continuum are assumed to be short, it must follow that when any point A' has been activated in the past and followed by activation of some structure X', other points (A_ns) on A will also have been aroused at at the same time, the probability of this simultaneous arousal decreasing with increase in the distance of the given A point from A'. As a consequence, structural relations between these A points and various Xs will have developed,[9] their distance being an inverse function of the distance A'–A_n. Though separate connections from A to X are drawn for

---

[9] See page 328 on restructuring and learning for the discussion of conditions of building a relation such as this.



$S' \cdots \overset{p}{>} X \left\{ \begin{matrix} p(S' \cdots A') \\ \times \\ p(A' \cdots X_n) \end{matrix} \right. \rightarrow$	.09	.45	.63	.72	.63	.45	.09		
$\leftarrow \left. \begin{matrix} p(S'' \cdots SA'') \\ \times \\ p(A'' \cdots X_n) \end{matrix} \right\} S'' \cdots \overset{p}{>} X$			.09	.45	.63	.72	.63	.45	.09
Combined Probability	.09	.45	.66	.85	.86	.85	.66	.45	.09

FIG. 2. Activation surface derived by combining probabilities of activation to each point on X from all related points on A. S' and S'' are stimuli activating points A' and A'' on continuum A. Various points are related to other points with the probabilities indicated below:

$S' \rightarrow A' = .9$
$S'' \rightarrow A'' = .9$
$A' \rightarrow X_4 = .8$
$A' \rightarrow X_6 = .8$
$A' \rightarrow X_3 = .7$
$A' \rightarrow X_5 = .7$

$A'' \rightarrow X_5 = .7$
$A'' \rightarrow X_7 = .7$
$A' \rightarrow X_2 = .5$
$A' \rightarrow X_6 = .5$
$A'' \rightarrow X_4 = .5$
$A'' \rightarrow X_8 = .5$

$A' \rightarrow X_1 = .1$
$A' \rightarrow X_7 = .1$
$A'' \rightarrow X_3 = .1$
$A'' \rightarrow X_9 = .1$

convenience, the more accurate picture would represent a sheaf or cone of relations without gaps in points on A that are connected with Xs.[10]

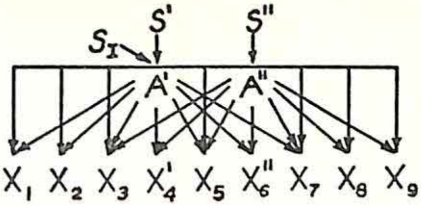
The nature of the activation surface at X will depend on the input to each X from all A sources, combined according to the assumptions of Principles II and III. The activation surface of Fig. 2, labelled *combined probabilities*, is obtained when the distances between the various points are assumed to be those indicated in the figure caption, when input to the two points on A are close enough together to provide multiple input to all Xs except those on the end of the continuum, and when the stated combining principles are applied to obtain the probabilities.[11] For example, in Fig. 2 in the row marked $p(S' \cdots > A') \times p(A' \cdots > X_n)$, the probability below X_1 (0.09) is the product of the probability that S' will activate A' (0.9) and that A' will activate X_1 (0.1). In the same row, the probability under X_2 (0.45) is the product of $p(S' \cdots > A')$ and $p(A' \cdots > X_2)$ and so on for this row and the next one which involves probabilities deriving from S'' activation. The row below the line (combined probabilities) is the combination of multiple input at each X point. For example, X_3 has a probability of arousal by S' equal to 0.63 and by S'', 0.09. The combination $[1 - (0.37 \times 0.91)]$ is 0.66 (Principle III).

Under these conditions, it is clear that high probability activation occurs at points between X' and X'', and that the probabilities fall off rapidly on either side of these maxima.

---

[10] We have not taken account of possible input to Xs from points on A other than A' and A''. If this were done, the probabilities of the intervening points would be even higher.

[11] Obviously such a surface will change not only with change in distance $A' - A''$, but also with different values of each distance in the structure. Also it should be noted that, as the probability values approach unity, less change will be produced by the existence of multiple input to any point.



$$\begin{array}{lcccccccccl}
\overset{p}{S'\cdots X}\left\{\begin{matrix}p(S'\cdots A')\\ \text{x}\\ p(A'\cdots X_n)\end{matrix}\right. \rightarrow & .09 & .45 & .63 & .72 & .63 & .45 & .09 & & & \\
\overset{p}{S_I\cdots X}\left\{\begin{matrix}p(S_I\cdots A')\\ \text{x}\\ p(A'\cdots X_n)\end{matrix}\right. \rightarrow & .08 & .40 & .56 & .64 & .56 & .40 & .08 & & & \\
 & & & .09 & .45 & .63 & .72 & .63 & .45 & .09 & \leftarrow \left.\begin{matrix}p(S''\cdots A'')\\ \text{x}\\ p(A''\cdots X_n)\end{matrix}\right\} \overset{p}{S''\cdots X} \\
\hline
 & .16 & .67 & .85 & .94 & .94 & .91 & .69 & .45 & .09 & \text{Combined } p
\end{array}$$

FIG. 3. Activation surface derived from combining probabilities of activation to each point on X when there is additional input to X from instructions (S_I) *to judge* S'.

Furthermore, the points between X' and X'' are activated with higher relative frequency than any others on the surface. This assimilation surface involves, then, maximum probability of activation at points intermediate between those positions that would be maximal if S' or S'' were presented and judged alone.

But a symmetrical surface of this kind results only if there are no sets, instructions, contexts, or other factors that provide differential input at A' and A''. Figure 3 suggests the effects of such differential input on the surface. Suppose the subject is told to judge the intensity of S'. It is suggested that such an instruction produces additional input to A', deriving from a motivation system set up by the instructions. This added input, identified as $S_I \cdots > A'$ in Fig. 3, is assumed to be equal to 0.8 and will have the effect of skewing the activation surface and of producing maximum probability of arousal at a point which is closer to X' than in the case of Fig. 2. Instructions *to judge* S'' would skew the surface in the opposite direction.

If the same principles of combining input at various points, the same distances within the A dimension and between A points and X points, and the same instructions to judge S' are assumed, we obtain the series of surfaces of Fig. 4 and 5 as the distance between the two points of input to A is increased. It will be seen that in

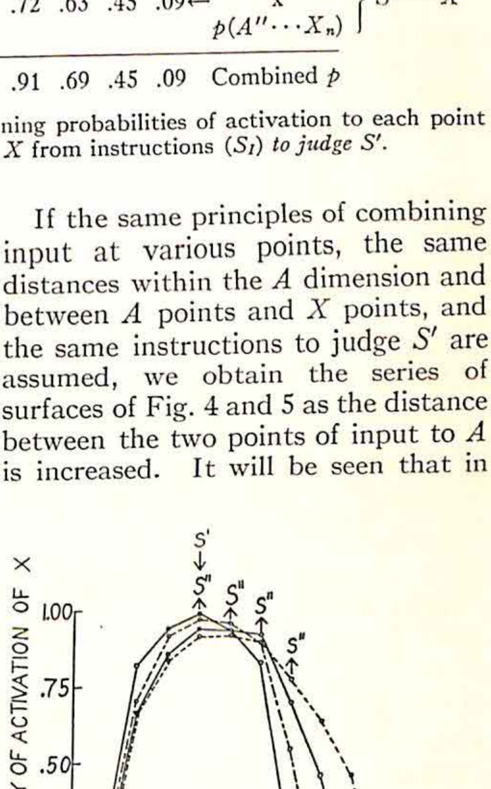


FIG. 4. Graphic representation of activation surfaces resulting when distances between points of activation are short. Surfaces A, B, C, and D. S' and S'' represent input points for the different surfaces.

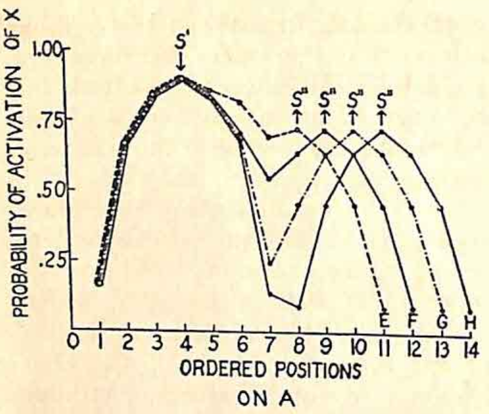


FIG. 5. Graphic representation of activation surface resulting when distances between points of input are greater. Surfaces *E*, *F*, *G*, and *H*.

passing from surface *A* to *H*, the breadth of the surface increases, bimodality increases, and the values of the probabilities of activation at intervening points decreases. The relative number of events occurring between points of activation must therefore increase because of increased distance and because adjacent points will not so often be activated simultaneously (Principle IV).

Change in the relative number of distinguished events as a function of distance between two points of activation. Figure 6 represents an estimate of the change in the relative number of events occurring between two input points with increase in the distance between the points, this distance being inferred from the difference in the judged positions of the points when each stimulus is presented and judged alone. In Fig. 6 the average probability of simultaneous activation of those adjacent pairs of points lying between input points is used as an index of the relative number of events (Principle IV). Operationally, the average probability that adjacent *X*s will be activated is assumed to be

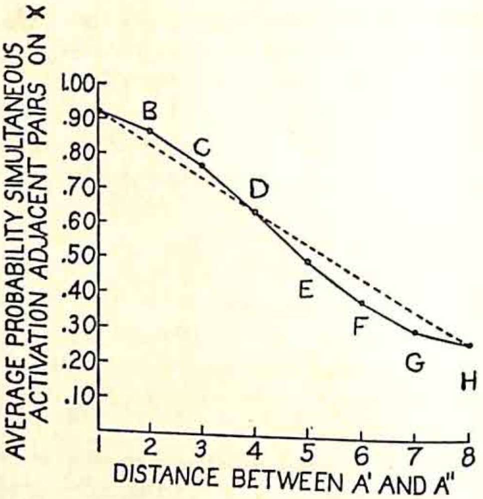


FIG. 6. Relation between difference in judges position of *S'* and *S''* when presented separately (distance *A' − A''* inferred) and the average probability of simultaneous activation of adjacent positions between input points for activation surfaces *A* through *H* of Fig. 4 and 5. Dotted line represents the relation of distance and average probability (inverse of relative number of intervening events) if the events increased directly with distance. The solid line results from plotting probabilities for the two-point activation surfaces.

equal to

$$\frac{\Sigma(p_{x'}\cdot p_{x'+1}) + (p_{x'+1}\cdot p_{x'+2})\cdots(p_{x''-1}\cdot p_{x}{''})}{N}$$

where $p_{x'}$ = probability of activation of x'; $p_{x'+1}$ = probability that X adjacent to X' is activated; $p_{x''-1}$ = probability that the X adjacent to X'' and lying between X' and X'' will be activated; and N = number of adjacent pairs between X' and X''.

This average probability of simultaneous activation was calculated for each surface, *A* through *H*, and the points joined by the solid line represent the average probability value for a given surface at the input distances from which that surface resulted (see Fig. 4 and 5). The dotted line joins the point derived from surface *A* (zero distance between input points) and the point derived from surface H (no overlapping input to intermediate X points). When the solid line is above the dotted one, this means there is a smaller average probability of simultaneous activation of adjacent points, and hence fewer intervening events for that surface than would be expected if the number of events were a linear function of input distance. If the solid line falls below the dotted one, there are relatively more intervening events for the surface.

It will be seen that the curve drops at an accelerated rate for surfaces B, C, and D, and that it falls at a decelerated rate for D, E, F, and G. This implies fewer events per A'–A'' unit of distance for the surfaces deriving from short distances between input points, and relatively more events per unit for the longer distance surfaces. Within the first group assimilation occurs; within the second, contrast,[12] for as a result of this disproportionate increase in the number of events at the longer distances, the judged position of the two stimuli will be shifted toward the ends of the continuum to produce the contrast effect.

The question inevitably arises how these activation surfaces come to produce a single response, such as the judgment of the position of S', which in the contrast condition will be displaced toward the end of the continuum. We must assume, without presenting the justification for the assumption, that this pattern of events (the activation surface) has become related to a structure representing some such response as *S' has low intensity of value E.* Another activation pattern consisting of two maxima and fewer intervening events would be more closely related to a point of the structure whose activation produces another judgment. But such responses will be complexly determined by input from points of maximum probability of activation on X combined with the process representing the number of intervening events. This suggests that contrast-determined responses should have longer latency than those resulting from assimilation.

Algebraic Summation and Cancellation Effects

It will be recalled that opposites structures involve continua which have developed subcategories as the distances along the continuum are increased. In considering the influence of such arrangements on the course of activation, we must ask first whether assimilation and contrast occur within such structures.

---

[12] Unfortunately, there is not space to discuss the relation of adaptation level theory to this formulation. In brief, adaptation level is conceived as that X point (X_{AL}) which has the largest number of high probability connections from points on continuum A. X_{AL} will be the point or points directly activated by the As most frequently aroused in the past, because as a result of this frequency X_{AL} will have had the best chance of becoming connected with all A points. If A has been activated at a number of points, those X positions near the center of the series will have the highest probabilities of arousal from any point on A, because such central points are closer on the average to other points on the continuum than are points at the end of the continuum. It is for this reason that the AL tends to lie at an average value of the experienced intensities of a stimulus. Given such a structure, the activation of any point A_n will elicit not only its own X_n with high probability but also X_{AL} through the generalized connections. In this way an activation surface is set up, the properties of which determine whether X_n and X_{AL} are assimilated or contrasted in exactly the same fashion that a surface produces assimilation or contrast when there are two *external* sources of input to A. The effect of this on the judgment of a single stimulus has been neglected for the sake of simplicity.

Take, for example, the opposites structure underlying positive and negative evaluation or affect. Continuum *A* is activated by input at two points. Suppose that the *X*s most readily elicited by activation on one side of the zero point of this continuum represent unpleasant affect, and on the other side, pleasant affect. The mid-category of this series will refer to zero affect, which is a definite position or structure whose activation produces this response. Insofar as there is generalized input from points along the positive-negative continuum to intermediate *X* structures including the zero category, assimilation or contrast will result for the same reasons and under the same conditions as they appeared within the simple similarity dimension.

But the result will appear as algebraic summation of the two activated points. Thus, if one activation point lies near the extreme negative end of the continuum and the other point activated at the same time is near zero but on the positive side, the intermediate point of maximum activation should be negative. Conversely, with high positive and moderately negative points activated, the result will be slightly positive. But if the activation points are equidistant from zero, then the zero category will be the one most probably activated. In other words, the phenomenon of algebraic summation of positive by negative affect is a special case of the more general theory of assimilation, and depends on (*a*) the existence of ordered arrangements of positions on a continuum, (*b*) input from the two points of activation to intermediate points, and (*c*) the existence of subcategories within the dimension such that different responses are produced if the point of maximum activation is on one or the other side of zero.

There is some evidence that such summation of positive and negative affect actually occurs. For example, Rosenberg's (**32**) successful prediction of attitudes toward segregation from the combination of positive and negative consequences related to segregation implies such summation. And Osgood and Tannenbaum's evidence for their congruity theory of attitude change also implies this same pooling of affect organized around a concept. Although they emphasize the assimilative aspect of the process, they do recognize that assimilation may not take place, when they say:

> The amount of incredulity produced when one object of judgment is associated with another by an assertion is a positively accelerated function of the amount of incongruity which exists and operates to decrease attitude change, completely eliminating change when maximal (**25**, p. 47).

In other words, "incredulity" tends to appear when the affect attached to a statement attributed to some source is distant from the level of affect (or evaluation) attached to the object about which the statement is alleged to have been made, so there is no assimilation.

Presumably contrast may also occur within an opposites structure when the distance is greater, provided the points activated are not at the ends of the continuum. Contrast is obviously impossible in this latter case, because activation cannot be shifted to more extreme points on the continuum even though the activation is of the sort that usually leads to contrast. The extreme response is given when the stimuli are presented alone; it cannot be any more extreme when the stimuli occur together.

Balanced and Unbalanced Structures

Heider (**11**), Newcomb (**23**) and others have used the language of Lewinian field theory in discussing "strains toward balance and symmetry." More recently, Cartwright and Harary (**7**) have described the same situations in the more general terms of graph theory. It is our hope that the same material may be handled ultimately by the concepts formulated in this paper.

Let us describe a structure which appears to have the properties of balance. A friend (*F*) has in the past often elicited in an experiencing individual strong positive feelings and on rare occasions negative feelings. Assume that these experiences have resulted in building the structural arrangements of Fig. 7. *F* is close to *P* (strong positive affect, $p = 0.9$); *F* is far from *N* (negative affect, $p = 0.1$). Positive and negative affect lie on an opposites dimension, the ends of which are represented as being far apart ($p = 0.1$). The relations are symmetrical.

On some occasion the friend is present as a stimulus which is assumed to provide cues for the direct activation of structures *F*, *N*, and *P*, the probability that *S* will activate each of these structures being equal to 0.9. What will the outcome be?

On the cycle following input to these points, *F* and *P* will each be activated with a probability of 0.83, while the probability of *N* will be 0.17. These values are derived by applying Principles II and III to the situation. For example, the probability of the activation of *P* from the direction of *F* is the product of the probability that *F* will be activated by *Sf* (0.9) and that *P* will be activated when *F* is active (0.9). This is equal to 0.81. The probability that *P* will be activated from the direction of *N* is the product of 0.1 and 0.9, and is equal to 0.09. Combining this multiple input (0.81 and 0.09) yields a probability of 0.83. In similar fashion, the probability of activation of *N* and *F* may be calculated for this cycle.

The obvious fact here is that *F* and *P* are activated with higher probabilities than *N*, and therefore will more often be aroused at the same time (p of simultaneous arousal $= p_F \times p_P$). As a consequence, *F* and *P* (the friend and positive affect) tend to be part of the same whole or the same unit event from which *N* is excluded, because of its distance from the other two structures and the lower probability of its being activated at the same time as they (Principle IV).

This structure is balanced, then, in the sense that (*a*) its activation will tend to result in two events or wholes with relatively clear-cut boundaries, *F-P* on the one hand and *N* on the other; and (*b*) the activation should be of short duration because the isolation of the *F-P* segment from *N* makes this essentially a simple two-

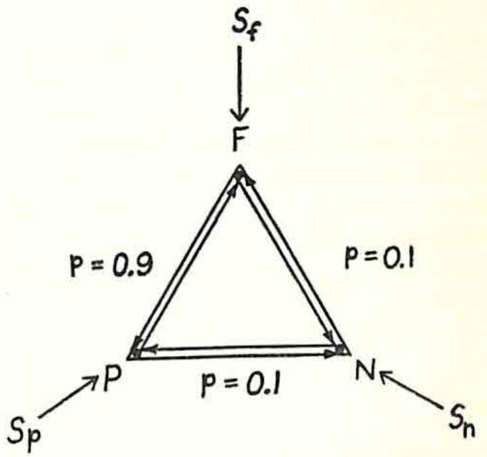


FIG. 7. A "balanced" structure. The *p*s are probabilities of activation of one point as a result of activation of the point connected with it by a line. Arrows indicate direction of the relation which is in each case symmetrical. *S*s are points of stimulus input.

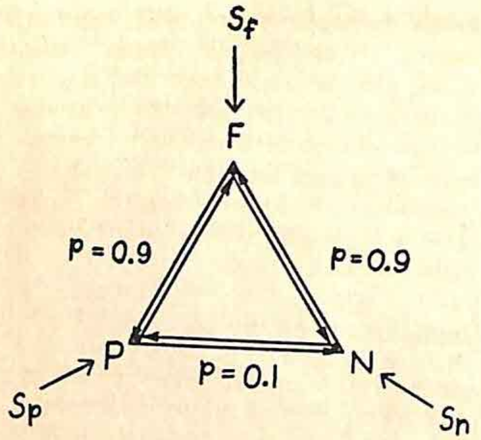


FIG. 8. An unbalanced structure.

part structure consisting of the *F-P* segment and the *N* segment. The mutual activation of *F* by *P* and *P* by *F* would be expected to produce rapid activity decrement because of the simplicity of the pattern and the short distances within this *F-P* segment. Activity within *N* should be of even shorter duration (Principle V).

Figure 8 represents certain aspects of a structure assumed to underlie strongly ambivalent attitudes toward *F*, and is offered as an illustration of an unbalanced arrangement. Its characteristic is that two events which are far apart (*P* and *N*) are both close to a third event (*F*). Thus the friend (*F*) is close to positive feelings but equally close to negative feelings, while *P* and *N* are again distant from each other. This time the activation probabilities on the first cycle are $F = 0.97$; $P = 0.83$, and $N = 0.83$. Given these values, the probabilities are high that these three points will be activated simultaneously and that they will be part of the same whole: $p(F \text{ and } P) = 0.8$; $p(N \text{ and } F) = 0.8$, and $p(P \text{ and } N) = 0.69$. However, despite the fact that the probability of the simultaneous occurrence of *P* and *N* is relatively high, they are distant from each other and so will not so readily become a unit event (Principle IV). The surfaces resulting from the activation of either *F* and *N* or *F* and *P* would be assimilative in pattern. However, any pattern which includes *F*, *N*, and *P* will inevitably lead to diffuse activation, since negative (*N*) and positive (*P*) affect will typically be associated with different structures. Moreover, since in this case, unlike the balanced structure, *N* and *P* are related through *F*, the surfaces, though diffuse, are not clearly separable. As a result, it would be expected that activity within this structure would tend to be maintained because distances are great enough to prevent early decrement (Principle V) but short enough (by way of *F*) to prevent early cessation of activation.

Heider's original discussion of balance described structures in which the terms were people, related by liking, disliking, similarity, causality, and so on. For example, he would regard as balanced a structure in which *A* dislikes *B* and *C*, while *B* likes *C*. This will be seen to have the same form as the structure of Fig. 7, provided these assumptions are made: (*a*) that the Heider situation represents a cognitive structure in some individual, and (*b*) that the relation of liking is converted to psychological distance between concepts of persons *A*, *B*, and *C*, and hence to the probability of activation of one term of the relation by activation of another one. Then, if disliking implies a distance between persons, it will be a greater distance than that implied by liking. Liked persons are closer to the liker than are disliked persons.[13] Here, as before, the structure is balanced in the sense that *B* and *C* act as a unit with consequent assimilation, while *A* is isolated. In Heider's unbalanced arrangement, *A* likes *B* and *C* but *C* and *B* dislike each other. Again, one event, *A*, is close to two others, *B* and *C*, which are far apart, and this pattern will have the properties of the schema of ambivalence in Fig. 8.

---

[13] It is interesting to note in this connection the common root in the phrases to *like*, to be *alike*, and to be *likely*. This suggests a common meaning for liking (positive attitude), being alike (similar or short distance

These examples suggest that the basic characteristic of the arrangements which have been called *balanced* is the fact that they provide the conditions for the development of focused and circumscribed activation surfaces which result in high probability activation of definite reactions and prompt completion of activity within the particular structure. On the other hand, the so-called *unbalanced* structure leads to a more diffuse activation surface and longer continued activation within the structure.[14]

It is of interest that Heider's unbalanced structure has been shown by Kogan and Tagiuri to be relatively rare (**18**). Thus configurations in which a person we like is conceived to choose one we dislike, or in which a person we dislike chooses one we like, are both less frequent than would be expected by chance. This should follow from the fact that in the unbalanced structure events such as these all tend to be activated simultaneously, frequently, and repeatedly with the consequence that the distances between parts will tend to become shorter, other things being equal. In other words, this unbalanced system of relations provides certain important conditions for restructuring and learning which leads to a new structure in which the parts are more closely related and hence tend to form a whole.

Motivation

It should be noted that those properties which lead to imbalance as defined above are the essential properties of a motive structure. The author has stated in an earlier paper (**27**) that when there is a certain degree of disparity between two events which are related or similar in some way, activation tends to be maintained in the system and thus to provide the continuing activity which is one characteristic of the motivated state. This activity continues until some change in the situation reduces the disparity or makes it so great that activity ceases. Thus disparity between the perception of one's status and that of his neighbor may constitute such a system, provided (*a*) that the two events, own status and that of the neighbor, are close enough together to produce a relatively high probability of activation of one term of the structure by activity in the other; and (*b*) that the disparity (or distance) and/or the number of parts are great enough to prevent activity decrement. Stated in terms of the present discussion, this amounts to saying that a motivated state exists when there is distance between events (disparity) great enough to prevent their being reacted to as a single whole with consequent activity decrement, and when at the same time the events are close enough together or related in such a way that disparate terms of the structure are mutually

---

apart), and being likely (high probability of activation of one by the other).

[14] There are other interesting properties of these structures that cannot be discussed now. For example, a greater average difference in the distances between the same two points via different paths may be an important characteristic of the unbalanced structure; that is, the distance from *A* to *B* may be much greater or less than distance *A* to *C* to *B* in unbalanced structures, but more nearly equal in balanced structures.

activating, perhaps by a common associate such as F in the unbalanced structure of Fig. 8.

SUMMARY

This paper has proposed that the course of events within psychological space is determined on any occasion by the nature of the structure, i.e., by the position of parts and the distance between them; and by the points in the structure which receive activation from some source at the moment in question.

Structures involving relations of similarity, opposition, complements, antecedents, and consequents are described in terms of the position of parts and the distance between them. The course of activation within certain of these structures is predicted on the basis of assumptions about the following: (*a*) the relation of distance to the probability of activation; (*b*) the combination of probabilities when there is multiple input to the same point in a structure; (*c*) the combination of probabilities when parts of a structure are arranged serially; (*d*) the number of events that are distinguished within an activation pattern; and (*e*) the development of activity decrement as a result of activity.

Suggestions are made about the way in which these principles may account for phenomena such as context effects, assimilation and contrast, algebraic summation of opposites, "balanced" and "unbalanced" functions, and the persistence of activity in the motivated state.

REFERENCES

1. Asch, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, 41, 258–290.
2. Atkinson, J. W., & McClelland, D. C. The projective expression of needs. II. The effect of different intensities of hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.
3. Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century, 1942.
4. Bridges, K. M. Emotional development in early infancy. *Child Develpm.*, 1932, **3**, 340.
5. Bridgman, P. W. *The logic of modern physics.* New York: Macmillan, 1932.
6. Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
7. Cartwright, D., & Harary, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63**, 277–293.
8. Coombs, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden Press, 1953. Pp. 470–535.
9. Freud, S. *Collected papers.* Vol. IV. London: Hogarth, 1952.
10. Heider, F. Social perception and phenomenal causality. *Psychol. Rev.*, 1944, **51**, 358–374.
11. Heider, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
12. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
13. Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 613–689.
14. Hovland, C. I., Harvey, O. J., & Sherif, M. Some basic processes underlying reaction to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.
15. Howes, D. H., & Osgood, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **6**, 241–258.
16. Hull, C. L. *Principles of behavior.* New York: D. Appleton Century, 1943.
17. Koffka, K. *Principles of gestalt psychology.* New York: Harcourt Brace, 1935.
18. Kogan, N., & Tagiuri, R. Interpersonal preference and cognitive organization. *J. abnorm. soc. Psychol.*, 1958, **56**, 113–116.
19. Köhler, W. *Dynamics in psychology.* New York: Liveright, 1940.

20. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems in social psychology.* New York: McGraw-Hill, 1948.
21. LEWIN, K. *Dynamic theory of personality.* New York: McGraw-Hill, 1935.
22. MEEHL, P., & MCCORQUODALE, K. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
23. NEWCOMB, T. M. Motivation in social behavior. In M. B. Jones (Ed.), *Current theory and research in motivation.* Lincoln: Univer. Nebraska Press, 1953.
24. OLDS, J. Physiological mechanisms of reward. In M. B. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. Nebraska Press, 1955.
25. OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.
26. PARKE, N. G., & SAMSON, E. W. Distance and equivalence in sequence space. Air Force Cambridge Res. Lab., June, 1951.
27. PEAK, HELEN. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. Nebraska Press, 1955.
28. PEAK, HELEN. Psychological structure and person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception.* Stanford, Calif.: Stanford Univer. Press, 1958.
29. PIAGET, J. *The origin of intelligence in the child.* London: Routledge & Kegan Paul, 1953.
30. POINCARÉ, H. *The foundations of science.* New York: Science Press, 1929.
31. ROBINSON, E. S. *Association theory today.* New York: Century, 1932.
32. ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.
33. SCHRÖDINGER, E. *What is life?* Cambridge: MacMillan, 1946.
34. SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.
35. STOUFFER, S., GUTTMAN, L., SUCHMAN, E. A., LAZARSFELD, P. F., STAR, S. A., & CLAUSEN, J. A. *Measurement and prediction.* Princeton: Princeton Univer. Press, 1950.
36. THURSTONE, L. L. Attitudes can be measured. *Amer. J. Sociol.*, 1928, **33**, 529–554.
37. WALKER, E. L. The course and duration of reaction decrement and the influence of reward. *J. comp. physiol. Psychol.*, 1956, **40**, 167–176.
38. WERNER, H. *Comparative psychology of mental development.* New York: Harper, 1940.
39. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received September 20, 1957)

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated.................................
Accs. No..............................

Psychological Review
Vol. 65, No. 6, 1958

BASIC ISSUES IN PERCEPTUAL THEORY

W. M. O'NEIL

University of Sydney

An attempt to classify modern theories of perception will be used to bring out some major points at issue between them. Some of these issues are epistemological but, as the psychologists' attempts to shelve such questions as nonpsychological have proved unavailing, it is better to face them without too much apology.

Five bases of classification are proposed. These interpenetrate yielding cells, as it were, in a five-dimensional solid. Though the five bases are logically independent, there is an historic tendency for the position taken on one to be associated with that taken on another. Thus some of the cells are recipients of a null class. This fact enables a simpler classification, which nevertheless is best approached through the more elaborate one. Incidentally, the scheme being proposed readily accommodates some other distinctions between types of perceptual theory offered by Koffka (27), Brunswik (12), Bruner and Postman (11), and Boring (7).

Perceiving as an active or as a passive process. The issue is whether perceiving is something done by or something done to the organism. Those who think of it as apprehension, or as the attainment of knowledge in the course of the organism's coping with its surroundings, take an activist position, whereas those who think of it in terms of the reception of stimulation, or as the input of information, are passivist. It is improbable that any contemporary theorist is as thoroughgoing a passivist as was, say, James Mill (35). With the common agreement that perception is selective, there is an admission that the organism is at work on the available stimuli or information. Nevertheless, there is a strong passivist tendency amongst the more behavioristically minded of the contemporary theorists. After Titchener there has not been any strong tendency of this sort amongst the phenomenalists. Within the activist theories there are some important differences. Primary among them is the treatment of perception as reactive or as proactive, to use Murray's terms (37). Those who regard perception as a discriminative response and who stress the impact of needs, motives and the like upon this response are reactivists. The more thoroughgoing purposivists, rarer in perceptual theory than in personality theory, it must be admitted, are proactivists.

The perceived as a real or as a phenomenal object. The issue, an essentially epistemological one, is whether we perceive a real or externally existing object or a representation (or some corresponding impression) of it. The man in the street, who is a naive realist, believes that when he sees a door, he does in fact see an object existing in the external world. A sophisticated phenomenalist such as Scheerer (42) maintains that a door existing in the external world reflects patterns of light energies which stimulate the retina, thus giving rise to neural processes which in turn produce a unit in the phenomenal field which represents the door. Though Scheerer and others of similar persuasion differ from Johannes Müller in many important ways, they agree to the principle in his statements: "Sensation consists in the sensorium receiving through the medium of the nerves, and as the result of the action of an external cause, a knowledge of certain qualities or conditions, not of external bodies, but of the nerves of sense themselves" (36, p. 1065) and "The immediate objects of the perception of our senses are merely particular states induced in the nerves, and felt as sensations" (36, p. 1073).

Early experimental psychology inherited a clear phenomenalist tradition from both sides of its family tree. Indeed, the analytic phenomenalism of British associationism blended quite readily with the analytic phenomenalism of the sense-physiologists such as Bell, Müller, and Helmholtz. The rise of realism in British philosophy in the early years of this century gave some special twists to the tradition, but it was American realism, especially working through behaviorism, which rejected it. However, the behaviorists largely vacated the field of perception. Their prime interest was in effector-based responses, and any interest they had in sensory processes was limited in the main to questions of sensitivity. There was an early promise, made first by the philosopher Perry (39), to give a full treatment of perception in terms of discriminative response, but only recently have serious efforts at fulfilment been evident. Thus it has been left largely to the phenomenalists to cultivate the greater part of this field.

The problem of error in perception or, put the other way, of veridical perception is a genuine one for psychology as well as for epistemology. All too often the psychologist with phenomenalist convictions regards it as not his concern to make out how the organism attains knowledge of the external world; this includes, of course, the scientist attaining knowledge about his subject matter. As a psychologist he is ready enough to confine the organism to an acquaintance with phenomena, yet as an organism theorizing about perception he presumes an acquaintance with both appearance and reality.

Two phenomenalists as different as Titchener and Koffka used the fact of error in perception to distinguish the arena of psychology from that of physical science. Titchener (45) analyzed visual and other illusions in order to illustrate his distinction between experience viewed as independent of and as dependent upon an experiencer. Koffka (27) made his distinction between the geographical and the behavioral environment the more convincing by his description of the dog bounding into a snow-filled crevasse which he took to be solid snow-covered earth. In general the phenomenalists make out a persuasive case that we perceive not real things but appearances, and that we respond to things not as they are but as they appear to be. These propositions, however, make it impossible to deal with or to recognize the fact of error which had so large a part in leading to them. If it were the case that we know only phenomena, there would be no way for us to get to know about error. Only if we truly know the real on some occasions can we tell that on others we have fallen into error. Within the realist position it is not difficult to show how the fact of error in perception can be recognised, though precisely which piece of knowledge is fallacious may be quite difficult to decide. Were realism to make out that such a decision was easy, it would be embarrassed by the fact that scientific progress is so difficult. In accordance with the formula about the whole truth and nothing but the truth, we may distinguish errors of omission and errors of commission. The former is quite easy for the realist to handle. For instance, in terms of perception as a discriminative response, it may be said that the organism has failed to respond to the object or that it has failed to respond differently to different objects. The error of commission is illustrated by the situation where we seem to have an object of such and such properties before us and yet where other, especially later, experience convinces us that there was no such thing confronting us. The realist, content with the discriminative response view as it is usually formulated, has no answer to the phenomenalist's question: "Where and what was the apparent object?"

The perceived as a term or as a proposition. The distinction being made is a logical and not a verbal one. A term is any property, such as what the word "red" means, or any set of properties, such as what the words "the red piece of paper" mean. A proposition is the predication of one term of another. The words "the red piece of paper is on a gray background" express or symbolize a proposition. In the behaviorist theories, as well as in most phenomenalist theories of perception, the perceived object is treated as a term. Thus the organism discriminates the red paper from its gray background or the perceiver experiences the field as a red paper figure on a gray ground. But if we consider perception as attaining knowledge, at any rate what James (26) meant by *knowledge about,* we need to think of what is known as having propositional form. Facts are not in the ordinary sense things or terms; they are not what can be expressed by means of nouns, adjectives, verbs, or adverbs alone. Facts are what we need sentences to talk about, and knowledge of a fact is judgmental. Those adopting the propositional view have been in recent times few in number and almost always no more than implicit adherents. There being so many who have treated the perceived object as a term, a much greater diversity amongst them can be expected.

Three major conceptions of the percept can be distinguished amongst the phenomenalists. At one end are to be placed those whom Boring (7) calls the *Leipsigers,* and at the other are his *Berliners.* The *Grazers,* kept company by Spearman and a large band of functionalists, are located somewhere in between. The Leipsigers, especially in the purified form of Titchener, agree with the radical behaviorists in being satisfied with simple discrete substantial terms. Titchener's sensations, despite his talk of context, are just as simple and discrete as Watson's stimuli, despite his talk of situation. Titchener had no way of including relationships within the experienced, and a behaviorist who conceives stimuli either as energy sources or as energies is equally handicapped . The earnest attempts of the Hullians to explain transposition phenomena in terms of stimulus generalization show the distaste for the inclusion of relationships among the stimuli. The Grazers found a place in the percept for relations, but did so by regarding them as imported into or imposed upon the situation by the perceiver. Both they and the Berliners agree in admitting relations as well as qualities in the percept. But the two disagree upon how relations get there. For the Berliners they are not the result of some "production process," but are as primary as the qualities related.

There was a distinct inclination amongst the *Akt* psychologists to adopt a propositional view. Brentano's distinction between the "lower" act of ideation and the "higher" act of judgment included a recognition that the object of the latter was propositional. Stout was quite explicit when he said:

> In general, it would seem that a complete object [of consciousness] can be adequately described in language not by isolated words but by propositions capable of being asserted, denied, doubted or assumed. . . . When we are said to perceive or know something this may be taken to express the fact that we perceive or know that something is the case (44, p. 100).

More recent formulations have not been as explicit. But the recent analyses by MacCorquodale and Meehl (31, 32) and Meehl and MacCorquodale (33) reveal that expectancy in Tolman's theorizing is propositional in form. Insofar as expectancy is equivalent to hypothesis, as used by Tolman and Brunswik (47), and as later adopted by Bruner (9) and Postman (40), all four theorists may be claimed to adopt a propositional view. Helson (23), in stating his adaptation-level theory as a general theory of perception, gives some signs of subscribing to this view too.

Descriptive versus abstractive modes of analysis. In offering a characterization of something in a way which distinguishes it from other things, we may resort either to descriptive or to abstractive terms. The former are features of the object actually before us when we inspect it, features that stand up for inspection, whereas the latter are not literally present in the form in which our analysis might suggest they are. The Wundtians recognized most of the time that the elementary sensations with which they characterized a given percept were in fact abstractions. A pure sensation, as they conceived it, is not something that we ever experience. Likewise, the dimensions of experience which have been derived by Boring (6) from the later Titchenerian doctrine of sensory attributes (46) are not components of experience, but the dimensions of a reference frame within which some particular experience may be located. On the other hand, features like figure-and-ground, contour, and gradient of texture are deemed by those referring to them to be the very material or structure of our experience as we have it. They are descriptive terms. Wundt, the early Titchener, and the Leipsigers generally preferred elemental abstractive analysis; the later Titchener, followed by Boring and Stevens and their associates, have adopted a dimensional abstractive analysis; David Katz, Rubin, the Berliners, Gibson and the phenomenologists generally have employed a descriptive and usually structural analysis.

Preferred location of causal conditions. In seeking to give a causal account of perception, the early theorists stressed proximal stimulus conditions. Later there was a considerable swing to distal stimulus conditions, as Brunswik (12, 14) pointed out. But both place the emphasis upon what Bruner and Postman (11) call autochthonous factors, whereas these two, in concert with many others in the last decade or two, have emphasized needs and other related "central" factors. There is, of course, another kind of "central" factor represented by the brain fields of Köhler and Wallach (28) and of Krech (29, 30), and by the cell assemblies of Hebb (22). The classification of perceptual theories contained in Allport's treatise (1) is based primarily upon the determinants stressed in the theory.

It is appropriate at this point to comment upon the unfortunate tendency of psychologists to identify causes with their effects, or to assimilate them. Werner and Wapner (48), in arguing that if both sensory and motor (tonic) processes interact they must be of the one kind, provide an example of assimilation. The widespread resort to dispositional concepts such as attitudes and sets usually involves an identification. No more is said about them than that they have specified observed effects and that they differ from one another in the effects they each have. Operational or dispositional statements are well enough, but the trouble arises when the need for anything more is not felt. It is well enough to identify opiates by means of their soporific effects, but it is misleading to identify them with those effects. An interesting example of this sort of identification may be found in the instructive controversy between Boring (7, 8) and Gibson (17). Boring, supposing for argument's sake our planet to be more liberally endowed with satellites, asks what we would be perceiving when we perceived the size of a moon. He is not a naive realist, and so he is not satisfied by the reply "Why, the size of the moon we happen to be looking at!" His own studies of the apparent size of our one and only pre-Sputnik moon led him to conclude that major determinants of such a perception are the actual distance, size, and elevation of the moon, the perceived distance of the moon, the actual elevation of the perceiver's regard, and the perceiver's attitude. He goes on, quite reasonably, to suppose that these values may be used as the parameters of some function, ø, which would be invariant when the perceived size of the moon is invariant. So far, to adopt a Boring phrase, this is good science. But it is probably only a confusion of cause and effect when he says ". . . then, in judging size, you are perceiving not object size, not retinal size, but ø. To discover the object of perception, you have to discover what function of the stimulus is invariant when the perception is invariant" (7, p. 146).

A THREEFOLD CLASSIFICATION

The historical association of views on these five basic issues has been such that a workable division into discrimination theories, phenomenalist theories, and judgmental theories is possible. No one of these is a compact cluster, so that in any general discussion of them some exceptions have to be made. The second cluster is indeed a clear doublet, the rather tough-minded abstractive theorists standing somewhat apart from the more tender-minded phenomenologists. In the sections below each of these types of theory will be characterized in general terms. Attention will be paid to the relative strengths and weaknesses of the discrimination and the phenomenalist theories. Finally, as part of an attempt to bring out what has usually been left implicit in judgmental theories, it will be argued that this type of theory affords a hope of our being able to retain the best of the other two worlds.

Discrimination Theories

Discrimination theorists are ordinarily slightly on the activist side of center. They are ordinarily realists in that the stimulus (wherever it is to be located, distally or proximally, and however it is to be analyzed) is real in some sense and not phenomenal. They are ordinarily given to abstractive analysis. The stimulus to which the response is made is a term rather than a proposition. There is a marked tendency to pack as much as possible into the stimulus, which is scarcely distinguished conceptually from the cause of the response. Graham (19), indeed, disposes of determinants such as sets and attitudes by referring them, at least in experimental settings, to instruction stimuli. Boring, whose theorizing is best thought of as moving from the tougher-minded phenomenalism towards discrimination theory, speaks of the perceiver's attitude as one of the stimulus properties which is a parameter of ø.

Apart from any disagreements on specific perceptual problems, a few general distinctions can be made amongst the proponents of discrimination theories. Those with the greater physiological interests tend to be more proximally focused, and to look for perceptual determinants largely in the sensory mechanism. Those who have been most influenced by modern positivism tend to have the more sophisticated views about the stimulus, to have a greater "focal length," and to favor dimensional analysis through their concern with stimulus continua. Those who are led to their problems by practical considerations—such as the need to establish the perceptibility of signs, signals, displays, and the like—tend to be empiricist in the rule-of-thumb sense.

This type of theory has many attractions. Foremost amongst these is the scope provided for rigorous theorizing and rigorous experimentation. Another, of less certain value, is the avoidance of the bothersome problem of introspection. As the aim of a perceptual experiment within this theoretical context is the testing of limits of discrimination and the seeking of invariant relations between stimuli and discriminative responses, it does not matter what the responses be, whether verbalization, key-pressing, or making matches; and it does not matter how they be established, whether by conditioning or by instruction stimuli. When a verbal response is used, it does not have to be treated as a report. To treat it as a report the investigator must concern himself with what it alleges and not with its own distinctive features as an event. Such a concern raises the nagging doubts about intersubjective corroboration to which the discrimination theorist seems prone. Their avoidance in this way does not, however, preclude the discrimination theorist from dealing with many problems that have often been deemed to be subjective. For instance, it can be argued that, in establishing a subjective or phenomenal scale of brightness or loudness, the subject is being calibrated as a different kind of instrument for application to some physical variable (*vide* Bergmann and Spence, 3).

The discrimination theorist also has the advantage of the phenomenalist in that he can offer some coherent account of error, as Perry (39) and other American realists have demonstrated. However, saying that when in error the organism has made the "wrong" response does not go far enough. It needs to be shown in what way that response is wrong. When the outcome of an animal's having made some response is the survival or the death of that animal, or even when the outcome is a shorter or longer period of delay in obtaining food or a mate, it may be said that the response made was right or wrong. However, there are occasions when an illusion contributes to survival or to a maximization of benefits, but is nonetheless an illusion. What distinguishes it as an illusion is that the seeming knowledge in it is false. This same limitation obtains when this type of theorist attempts an account of knowledge in the sense of our store of knowledge. Apart from the limitation of the dispositional terms employed, for instance, by Hull (25) and later Berlyne (5), no way is provided for a distinction between true knowledge and false knowledge.

The advantages gained by the eschewal of introspection must have offset against them some handicaps accruing at the same time. It would appear helpful to know that chromatic colors disappear in scotopia, and that it is not merely a matter of some one chromatic color replacing all other colors. A purely discriminative experiment could tell us no more than that, apart from any intensity differences, stimuli of one wave length are not discriminable from those of another including a mixture of all wave lengths. We could have no indication whether the central processes following different stimuli were the same or whether these processes were functionally equivalent though qualitatively different. The pattern of verbal responses made with the addition of appropriate instruction stimuli would not afford a clue to this unless we regard such responses as alleging what is the case, i.e., as being introspective reports.

Finally, a discrimination theory usually involves a renunciation of what seems to some the genuine and important psychological question expressed by Koffka (27) as: "Why do things look as they do?" This question seems possible only if we are willing to consider introspective data and make some form of distinction between the real and the phenomenal. A distinction between tape-measured and subject-measured distance, for instance, does not go far enough. But as soon as the discrimination theorist begins to make the necessary admissions he is in danger of losing many of the advantages he had otherwise gained for himself. Berlyne suggests: "If we think of the perceptions which constitute the 'phenomenal world,' etc., as reactions to stimulation from physical energy-changes, and overt behavior in its turn as a reaction to them, we may be able to enjoy some of the benefits of both points of view" (4, p. 145). However, if perception is taken as a response in the ordinary Hullian sense, there are difficulties arising through there being no effectors to perform it; and if it is taken as a stimulus-producing response or purely theoretical or intervening variable, it is puzzling how we come to know that it has the properties attributed in other theories to the "phenomenal" world.

Phenomenalist Theories

Phenomenalists have a broader spread over the active-passive variable. Contemporary representatives more often than not come down on the activist side, sometimes well out towards the extreme. They are never realists in predominant flavor, although many struggle, as it were, towards realism. Generally speaking, the more distally focused have the stronger realist tendency. In the past they preferred elemental analysis, but now the predominant preference is for descriptive structural analysis. Once the basic classifying and counting had been done by the elementarists, the remaining part of the program appeared tedious and sterile. Phenomenology has afforded a much richer alternative. In both varieties, phenomena are conceived as terms—sensations when the focus is proximal, and objects when it is distal. The favored determinants are stimulus properties (though played down as compared with the discrimination theorists), other phenomena, brain states, and, with the more markedly activist, motives. There is some dispute, however, amongst them as to the legitimacy or profit in using one or other of these. For instance, Gibson (15) seems mainly concerned to find stimulus correlates and little impressed with the need to locate neural determinants, whereas Werner and Wapner (48) center their attention upon the total organismic state. The rocks that have seriously split phenomenalism have been the issue of abstractive versus descriptive analysis and the issue of proximal versus distal focus. Though independent issues, those choosing abstractive analysis have usually chosen a proximal focus, and those choosing descriptive analysis have usually chosen a distal focus.

Though an epistemological history of phenomenalism should go back at least as far as Locke's distinction between primary and secondary qualities, a psychological history can appropriately begin with that part of Müller's doctrine of the specific energies of the sensory nerves which stated that we are aware not of external objects, but of the states they produce in our sensory nerves. Within this doctrine a veridical knowledge of external objects would be attained to the extent that there is some parallelism or isomorphism between the external objects and the states they produce. When that parallelism breaks down we have error or illusory perception. But how anything about the parallelism or its absence can be discovered is a puzzle. It is not enough to say, as Müller did, that vision is ordinarily elicited by light, and that when it is elicited by pressure on the eyeball there is an incongruity of stimulus and sensation. To say this assumes that we may know light and pressure on the eyeball as other than phenomena. But even were there two sorts of knowledge—the sort we have in experiencing color and the sort we have in knowing that light is impinging on the eye—it is difficult to show how we would relate one to the other and to say in which realm of knowledge this relating occurred (*vide* Anderson, 2).

There is an interesting paradox in the phenomenalism of Müller's sort. Our knowledge of our sensory states must in one sense be veridical, i.e., in terms of the doctrine, we know only these states and can not know other than what is present in them. Yet it almost always seems to be knowledge of external objects, which, according to the doctrine, is a nonveridical impression. Further, what is said to be knowledge of the states of our sensory nerves gives us the knowledge we have of external objects and very little, if any, information about the states of our sensory nerves. The way we are able to get such knowledge is to have those states as the external objects of other sensory states. Müller, in effect, told us that it is not the moon that we see but a state of affairs in the visual mechanism, whereas Boring tells us that it is not the moon that we see but a mathematical function. Both seem to be confusing what we see with the conditions of our seeing.

Though the later phenomenologists reject sensory states located at the organismic surface, their objects are not literally *out there.* Their focus, as Brunswik (14) shows, is not fully distal. The term "intermediate object," which Brunswik used earlier in his work on object constancy, might well be applied with a slightly different meaning to the phenomenal object in these theories. There remains, nevertheless, the impassable gap between the phenomenal and the real. However, the phenomenological approach has the virtue of producing an orderly account of how things look. The now classical work of Katz, Rubin, and Wertheimer and the more recent work of Michotte (34) and Gibson (16) is evidence enough of this. Furthermore, such work has produced data which seem amenable to corroboration and which set what has already been said to be a genuine psychological problem.

Judgmental Theories

Perhaps some form of discrimination theory has greatest appeal to the physiologically minded, and some form of phenomenalism to the psychologically minded, especially when in the ranks of the professional physiologists. Judgmental theories have certainly been propounded most often by the philosophically minded. However, if, as will be asserted here, such a theory can accommodate some distinction of the sort attempted in the contrast between the real and the phenomenal, and at the same time preserve some of the advantages of objective reference attained by discrimination theories, it may well be the type of theory the psychologist will find best adapted to his needs.

Unfortunately, we have no contemporary judgmental theory in sufficiently elaborated form to let us prove these hopes. Bruner (9) and Postman (40) gave separate statements of their joint thinking which constitutes one promising outline. However, much that is relevant to the claims being made is left implicit in their theorizing. An explication is rendered the more difficult by a more recent paper by Postman (41) which reveals a distinct swing away towards a discrimination theory. A later paper by Bruner (10) extends some aspects of the judgmental treatment but does not relate its terminology to that of the earlier statements. Without asking these theorists to accept any responsibility for what is to be maintained, many things which perhaps they thought best to leave implicit will be made explicit. Then, in the light of that interpretation, some indication will be given of where their theorizing seems to have gone off (or not far enough along) the track favored here. Finally, some promissory notes will be proffered in the absence of hard cash to capitalize the expansion of this type of theory. Bruner, Postman, and their associates have already a great deal of hard cash to the credit of other aspects of their theorizing.

The full course of cognition was said in 1951 to be an hypothesis-information-testing cycle. This conception applied not only to perception but also to thinking such as remembering, imagining, and reasoning. This is an immediate attraction. All too often we are asked to conceive perception in one way and thinking, with all it entails, in so unrelated a way that we remain puzzled as to how what we acquire through perception can be incorporated in our thought.

The central notion is "hypothesis," which in Bruner's later statement seems to be replaced by "category." Hypotheses "serve to select, organize and transform the stimulus information" (40, p. 249); or "stimulus inputs may be sorted, given identity, and given more elaborated, connotative meaning" in terms of categories (10, p. 148). Hypotheses are deemed to vary in strength, or categories to vary in accessibility. But what hypotheses or categories are is a matter left a little obscure. Bruner in general was and is content with a dispositional conception—these are we-know-not-whats recognizable only by their determinants or by their effects. Postman went a little further in saying that an hypothesis is an expectation that such and such is so and so, or that such and such does so and so. He relates this to Tolman's notion of expectancy. MacCorquodale and Meehl have argued convincingly that an expectancy in Tolmanian theory has the form "if (S_1 and R), then S_2"; e.g., "if, when the light flashes, the lever is pressed, then a pellet of food falls into the cup." This, of course, is an action hypothesis which is the sort required in a theory of performance. Where the account puts the emphasis upon cognition, the "response component" may be ignored; thus "if S_1, then S_2." Cognition of whatever type is to be regarded then as a predicating (asserting, denying, supposing, doubting and so on) S_2 of S_1; or, if no imputation of conscious deliberation is made, as a judging of S_1 to be S_2. It is important to stress that S_1 and S_2 may be quite complex, and that a distinction between a judgment and the verbal or other expression of it is intended.

The judgment or predication may be true or false, depending upon whether, in fact, occasions of S_1 are occasions of S_2. In illusion there is a false predication. This arises in several ways, one of which is fallacious inference, though the latter is more prevalent in thinking than in perceiving. In an error of omission S_1 is judged to be S_2 alone when in fact S_1 is also S_3, S_4, etc. Thus we must recognise that there are, first of all and independently of our knowledge of them, the facts. These are the empirical content of the term "real" in the traditional distinction. When we judge that S_1 is S_2 and in fact it is, we have attained a knowledge of that fact. When we have attained a knowledge of some of the infinite facts confronting us and have failed, by omission and by commission, to attain the knowledge of others, we have the state of affairs emphasized by the term "phenomenal." To say that we know only the phenomenal is true in the trivial sense that we know only what we know. But it should not be interpreted as having more significance than that. It is demonstrable that, in a rough manner of speaking, the moon looks smaller than it really is. Further, if we assume, as some of our forebears did, that the law of visual angle is sufficient for the explanation of size perception, it may be said, equally roughly, that the moon looks bigger than it ought to. To ask, "What is this too small and too large thing that we see?" is to ask the wrong question. It assumes that there is such a too small and too large thing somewhere. As it is not *out there,* it is then concluded that in some sense it must be *in here* or somewhere in between. The question, "What do we see when we see the size of the moon?" is answerable properly not with "the real moon" (for that gives us no way of coping with the anomaly we have noted), nor with "the retinal image of the moon" (for that is contradicted by experience), nor with "an intermediate or a phenomenal object" (for that is a piece of fabrication which leads us into confusion), but with "that the moon is size X," or "that the moon is the size of X." That this visual judgment is false, and that it varies with the position of the eyes in the head, are matters of great interest to the psychologist—but not because he studies phenomenal moons, leaving the astronomer to take care of the real ones.

Frequently sense perception is regarded as having the particular and the concrete as its content whereas thought, apart from its purely reproductive forms, is regarded as being characteristically concerned with the universal, the conceptual, and the abstract. The psychologists prone to make this distinction seem to have forgotten the long-standing arguments about the nature of universals, and not to have learned that the scholastic positions known as realism, nominalism, and conceptualism do not exhaust the possibilities. Discrimination theorists, when dealing with cognitive matters other than perception, are prone to resort to a sort of nominalism: knowledge of universals is merely making the same response, a symbolic response, to several stimuli. Phenomenalists of the elementarist kind tend to follow a similar line, substituting such phenomena as composite images. Phenomenologists in psychology turn regularly to conceptualism, Goldstein and Scheerer (18) affording an example singularly rich with scholastic vestiges. Bruner, in his theorizing about categories, reveals an inclination towards conceptualism. Nevertheless, his discussion has the merit of showing that perception is not confined to the particular, and that a concern with the universal is as characteristic of perception as it is of thought. A clearer recognition that all cognition is judgmental would lead to a rejection of both nominalism and conceptualism without forcing one into the mystification of medieval realism. What distinguishes the universal is that it is predicable of many things, a point recognized by Bruner in respect of his categories. However, this is the case with anything that is predicable, even when it is a so-called concrete particular. Some psychologists seem to think that there are some logical considerations requiring a special class of "things" called concepts. The nominalists know that there are no such considerations, but the psychologists among them fail to see that speaking of stimulus and response generalization does not take them far enough in accounting for knowledge. It is not that we confuse one member with another of its class, but that we know each of them to be things of the same sort.

The explication given here of the term "hypothesis" (or "category") reveals a difficulty in giving a qualitative account of hypothesis-strength (or category accessibility) which has an important place in the theory. Though the operational criteria provided especially by Postman suffice for the mounting of an extensive and fruitful experimental program, they are not adequate for a full explanatory use of the term. An hypothesis itself cannot vary in strength; S_1 is simply S_2 but never more or less S_2. No ground is gained by saying that the hypothesis is held more strongly unless it can be said what holding an hypothesis is. To say that the hypothesis is used more or less frequently in certain situations does not explain what hypothesis-strength was meant to explain. The same strictures may be made with respect to habit-strength in Hullian theory. Should habits, now defined dispositionally, turn out to be neural structures, then habit-strength could be some variable property of those structures. A judgment may have a similar neural determinant, in which case hypothesis-strength could be given positive meaning. This is a hope, however, with little factual backing at the moment.

The information and test phases in the cycle also need some clarification. At first sight the information phase is a good deal like what we have all been calling perception. Were further consideration to confirm this impression, then this theory would be not so much an account of perception itself as of the wide context of processes in which it is embedded. Further, too much stress is placed by Bruner and by Postman on stimulation as such, that is, the impact of stimulus energy upon the receptor mechanism. It appears important to recall at this point the distinction made by Harper and Boring (21) between cues and clues. What is referred to as information should on some occasions be recognized as a nonperceived but sensorily mediated determinant of perception, and on others as a hypothesis (judged state of affairs) relevant through inferential processes to the hypothesis under test. The former type of information is a cue (which is all that is considered in cybernetic theory), whereas the latter is a clue which may be a premise as well as a mere determinant. It may well be that a similar distinction should be made in respect of the test phase of the cycle, yielding something like reinforcement in that brand of learning theory with which discrimination theories are most readily affiliated, and something like inference in Tolman's theory, which is at least implicitly judgmental. The strengthening of an hypothesis is akin to acquisition in learning, and the weakening of an hypothesis is akin to inhibition and extinction.

The distinction between cue and clue may well help with the resolution of the present paradox of size and shape perception. The bidimensional retina can of itself encompass only visual angle and projected shape. Perception of size should thus be dependent upon access to information about distance, and perception of shape upon access to information about inclination. There is a fair amount of evidence to support the former of these suggestions, though little available that is relevant to the latter. Several theorists, e.g., Koffka (27) and Boring (7), have been led to suppose that there must be an invariant relation, to put it in Boring's way, between, in the one case, perceived size, perceived distance, and retinal image size. Now an experiment like that of Holway and Boring (24) does not bear on this alleged relation, for no distance judgment was obtained. An experiment like that of Gruber (20) does bear on it; that it falsifies the alleged relationship is ameliorated only by its inherent design defects. Stavrianos (43) failed, despite a commendable persistence, to show that errors in perceived shape were related to perceived inclination. Her experiment, too, has some basic design defects. Nixon (38), in a series of studies free of these limitations, has falsified the hypothesized relationship. The puzzle here would be resolved were it to turn out that this is an "overdetermined" situation in the sense that there is more information, not all of which is congruent, than the minimum required for a given judgment of size or shape. Some of this information is in the form of a judgment of distance or of inclination. This is a clue to size or to shape when it is conjoined with information such as retinal size or retinal shape. It may operate as a clue through some inferential process, as Helmholtz would have maintained. In addition, there may be cues to distance (inclination is a matter of relative distances) which on this occasion do not influence the judgment of distance or of inclination, but which may influence the judgment of size or of shape. If ocular torsion is relevant to the moon illusion, it would be just such a cue: the paradox that Gruber leaves was left also by Boring in respect of that illusion, namely, that the changed apparent size of the moon is due to a changed cue to distance, even though the apparent distance has been inversely affected. The way a cue affects a judgment would not be through some inferential process. Inference seems peculiar to discourse or to knowledge. It is not merely a matter of cause and effect. When we speak of cues, no more than a causal relation is being referred to. The nearness of the horizon moon is possibly the result (through inference) of its appearing so large.

Were a considerably modified notion of response developed, one able to embrace what Woodworth (49) called a "mental" response as well as what he called a "motor" response, the proposal made by Berlyne (4) would have greater merit. As was pointed out above, a pure stimulus act such as he had in mind will not suffice. Nor, of course, is it enough to say that the perceptual response is to be characterized as discriminative. Little responding, whether cognitive or otherwise, is indiscriminate. The kind of response needed is judging that S_1 is S_2. To fit such a concept on to Hullian theory requires rather more general renovations than Berlyne suggests. However, if it can be done, we should have, as he is hoping to have, the best of the two worlds of discrimination theory and phenomenalism in coping with perceptual problems.

REFERENCES

1. Allport, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
2. Anderson, J. Empiricism. *Aust. J. Psychol. Philos.*, 1927, **5**, 241–254.
3. Bergmann, G., & Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–24.
4. Berlyne, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
5. Berlyne, D. E. Knowledge and stimulus-response psychology. *Psychol. Rev.*, 1954, **61**, 245–254.
6. Boring, E. G. *The physical dimensions of consciousness.* New York: Century, 1933.

7. Boring, E. G. Visual perception as invariance. *Psychol. Rev.*, 1952, **59**, 141–148.

8. Boring, E. G. The Gibsonian visual field. *Psychol. Rev.*, 1952, **59**, 246–247.

9. Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsay (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951.

10. Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

11. Bruner, J. S., & Postman, L. Perception, cognition and behavior. *J. Pers.* 1949, **18**, 14–31.

12. Brunswik, E. The conceptual focus of some psychological systems. *J. Unified Sci.*, 1939, **8**, 36–49.

13. Brunswik, E. Psychology in terms of objects. Reprinted in M. Marx (Ed.), *Psychological theory*. New York: Macmillan, 1951.

14. Brunswik, E. The conceptual framework of psychology. *Internat. Encyc. Unified Sci.*, Vol. I, No. 10. Chicago: Univer. Chicago Press, 1952.

15. Gibson, J. J. Studying perceptual phenomena. In T. G. Andrews (Ed.), *Methods of psychology*. New York: Wiley, 1948.

16. Gibson, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.

17. Gibson, J. J. The visual field and the visual world: A reply to Professor Boring. *Psychol. Rev.*, 1952, **59**, 149–151.

18. Goldstein, K., & Scheerer, M. Abstract and concrete behavior. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).

19. Graham, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.

20. Gruber, H. E. Relation of perceived size to perceived distance. *Amer. J. Psychol.*, 1958, **67**, 411–426.

21. Harper, R. S., & Boring, E. G. Cues. *Amer. J. Psychol.*, 1948, **41**, 119–123.

22. Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.

23. Helson, J. Perception. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand, 1951.

24. Holway, A. H., & Boring, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.

25. Hull, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511–525.

26. James, W. *Principles of psychology*. London: Macmillan, 1890.

27. Koffka, K. *Principles of gestalt psychology*. London: Kegan Paul, 1935.

28. Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. philos. Soc.*, 1944, **88**, 269–357.

29. Krech, D. Dynamic systems, psychological fields and hypothetical constructs. *Psychol. Rev.*, 1950, **57**, 283–290.

30. Krech, D. Dynamic systems as open neurological systems. *Psychol. Rev.*, 1950, **57**, 345–361.

31. MacCorquodale, K., & Meehl, P. E. Preliminary suggestions as to the formulation of expectancy theory. *Psychol. Rev.*, 1953, **60**, 55–63.

32. MacCorquodale, K., & Meehl, P. E. Edward Chace Tolman. In W. K. Estes et al. (Eds.), *Modern learning theory*. New York: Appleton-Century-Crofts, 1955.

33. Meehl, P. E., & MacCorquodale, K. Some methodological comments concerning expectancy theory. *Psychol. Rev.*, 1951, **58**, 230–233.

34. Michotte, A. *La perception de la causalité*. Louvain, 1946.

35. Mill, J. *Analysis of the phenomena of the human mind*. (2nd ed.) London: Longmans, Green, Reader & Dyer, 1878.

36. Müller, J. *Elements of physiology*. Vol. II. (Trans. by W. Baly.) London: Taylor & Walker, 1842.

37. Murray, H. A. Towards a classification of interaction. In T. Parsons & E. Shils (Eds.), *Towards a general theory of action*. Cambridge: Harvard Univer. Press, 1951.

38. Nixon, M. C. *Perceived shape and slant*. Unpublished thesis, Univer. of Sydney, 1958.

39. Perry, R. B. *Present philosophical tendencies*. New York: Longmans, Green, 1912.

40. Postman, L. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951.

41. Postman, L. Perception, motivation and behavior. *J. Pers.*, 1953, **22**, 17–31.

42. Scheerer, M. Cognitive theory. In G.

Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.

43. STAVRIANOS, B. K. The relation of shape perception to explicit judgments of inclination. *Arch. Psychol., N. Y.,* 1945, No. 296.

44. STOUT, G. F. *Manual of psychology.* (4th ed.) London: Univer. Tutorial Press, 1929.

45. TITCHENER, E. B. *A text-book of psychology.* New York: Macmillan, 1909.

46. TITCHENER, E. B. *Systematic psychology: Prolegomena.* New York: Macmillan, 1929.

47. TOLMAN, E. C., & BRUNSWIK, E. The organism and the causal texture of the environment. *Psychol. Rev.,* 1933, **42**, 43–77.

48. WERNER, H., & WAPNER, S. Toward a general theory of perception. *Psychol. Rev.,* 1952, **59**, 324–338.

49. WOODWORTH, R. S. A revision of imageless thought. *Psychol. Rev.* 1915, **22**, 1–27.

(Received November 27, 1957)

Psychological Review
Vol. 65, No. 6, 1958

A CONCEPT–FORMATION APPROACH TO ATTITUDE ACQUISITION

RAMON J. RHINE

System Development Corporation

The purpose of this paper is to attempt a theoretical formulation of attitude development from a concept-formation point of view. Mediating responses have been employed by Osgood (21) to explain concept formation, and a mediating or implicit response has also been used as a main theoretical construct in Doob's (8) outline of attitude theory. To this extent, there is already an established theoretical link between concepts and attitudes.

Doob calls attitude "an implicit drive producing response considered socially significant in the individual's society" (8, p. 136). This viewpoint has been criticized by Chein for a number of reasons, including the fact that it does not clearly indicate the evaluative nature of attitudes. Chein would rather call attitude a "disposition to evaluate certain objects, actions and situations in certain ways" (3, p. 177). Although many researchers in the field of attitudes might concur with Chein on this last point, a single meaning of attitude upon which there is close agreement is not available. Nelson's (20) review of the literature disclosed 23 more or less different definitions of attitude, and more recent statements could expand the list even further. No one definition seems clearly superior solely on logical grounds; if it were otherwise, there would not be so many alternatives.

The various definitions name constructs, and presumably some constructs have more scientific value than others. The psychologist, seeking a guide for his research among many alternatives, looks for a construct with heuristic and predictive value, particularly a construct which leads to informative research. Insofar as attitude change is concerned, Doob's type of approach has stimulated considerable research. But in the area of attitude development, very little carefully controlled experimentation has been done. Besides the research of Eisman (9), there is little experimentation available on attitude acquisition, although studies such as those by Horowitz (13) and by Clark and Clark (4) provide information about attitude growth. If it is possible to explain attitude learning in terms of concept formation, then attitude development is opened to controlled experimentation through concept-formation methods. A concept-formation approach to attitude learning also enables any systematic viewpoint which can account for concept formation to account at the same time for attitude formation.

THE GENERAL MEANING OF AN ATTITUDE

A concept is sometimes thought of as a psychological mechanism that represents a set of stimulus patterns; that is, a concept is considered a mental principle through which an individual can classify a number of objects in his stimulus world. Thus, redwood, elm, pine, fir, and oak fall into the class of "tall, nonanimal, living things," and *tree* is the name given to this principle. Similarly, the concept "a person who speaks with a drawl and has dark skin, thick lips, and kinky hair" classifies American Negroes for some people. Of course, the concepts learned in real life are not always verbalized into a precise principle. In fact, behavior in laboratory studies (19) indicates that a concept can be learned even though *S* is unaware of it and is unable to verbalize the principle on which the concept is based. The same phenomenon probably occurs with attitudes. People often betray by their behavior attitudes which they do not recognize as their own.

On a common-sense level, concepts are developed over a period of time through a series of experiences. The child learns the concept *cow* because his parents and others point to objects in the field or in pictures and call them by that name. He learns that color alone is not the key to cow-ness, but that it is helpful. All black objects are not called cow, but green objects are never called by that name. The child soon learns to discriminate between cows and other objects as he learns that cow-ness is a feature of not one stimulus pattern but of a whole class of patterns.

The child learns about people in much the same manner. He discovers that one dimension of Negro-ness is dark skin, but that a deep suntan does not make one a Negro. He gradually learns that Negro-ness involves other dimensions such as facial characteristics and hair properties. This learning eventually enables him to make fairly accurate discriminations between Negroes and other persons. So far, the child has a concept of a Negro but not an attitude, because an attitude involves an evaluative dimension. Suppose, however, he has some unpleasant personal experiences with Negroes or receives adverse information about them from his parents. Then the child's concept might be based on the dimensions, "dark skin, kinky hair, thick lips, speaks with a drawl, and *bad*." Since an evaluative dimension now appears in the concept, he has acquired an attitude.

This approach to attitudes is not an arbitrary substitute for other meanings of attitude, but is representative of commonly used meanings, as may be seen from the following quotations:

"Most researchers in attitudes agree that one of the functions of attitudes is to integrate social perceptions and experiences. Thus our experiences with our mothers, sisters, girl friends and maiden aunts are integrated . . . into an attitude toward women" (16, p. 292). ". . . attitude [is] a more or less permanently enduring state of readiness of mental organization which predisposes an individual to react in a characteristic way to any object or situation with which it is related" (1, p. 13). "The attitude is a way of conceiving an object; it is the mental counterpart of an object" (10, p. 11). ". . . the *meaning* of a concept is its location in a space defined by some number of factors or dimensions, and *attitude* toward a concept is its projection onto one of these dimensions defined as 'evaluative' " (22, p. 42). "An attitude, roughly, is a residuum of experience, by which further activity is conditioned and controlled . . . an inner mental organization takes place which predisposes the person to a certain type of activity towards objects, persons, and situations" (17, p. 238).

The thread which runs through these quotations—"experiences integrated into an attitude," "mental organization," "mental counterpart," "a way of conceiving"—is that an attitude is a kind of mental organization that is built through many experiences. In a sense, as a result of varied contacts with a class of objects, a concentrated representation of a set of stimuli is left within the person, which is the essence of what is generally meant by a concept.

AN EXPLANATION OF ATTITUDE

If an attitude is taken to be a concept with an evaluative dimension, its meaning may be derived from an explanation of concept formation. Consequently, any theory which can explain concept formation can also help to explain attitude development. One of several possible explanations of concept formation is provided by S–R Theory.

Borrowing from Osgood (21), a concept may be defined as the associations between a common response (often verbal) and a set of stimuli. These stimuli frequently comprise a class of phenomena which display certain common characteristics. The common characteristic of the set of phenomena may be the identical elements with which Hull (14) worked, such as a common shape embedded in each of several different Chinese characters. Or, as Smoke (30) has demonstrated, it may be common perceptual relations when, for instance, *S*s view a series of different figures and classify together all those having perpendicular lines in them. Although the stimuli mentioned in the definition of a concept often fall into a class, Osgood suggests that the only essential condition for concept formation is the associations between a common response and a variety of stimuli. There are no obvious stimulus characteristics common to hour-glass, ruler and mental test, all of which are *measuring instruments;* or to file cabinet, typewriter, and desk, all of which are *office equipment*. These items have in common the overt response of naming them, and, of greater theoretical importance, a common mediating process.

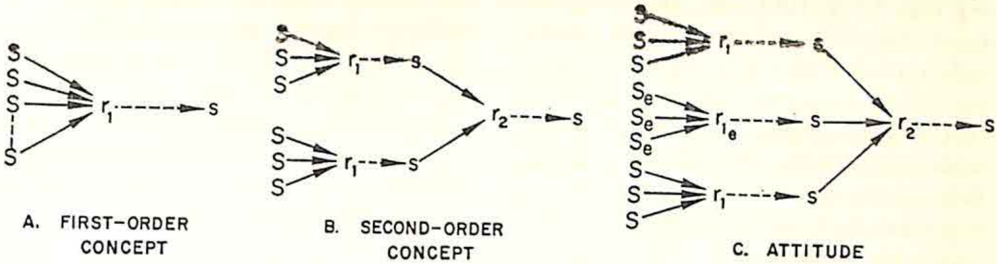
The mediating response, Hull's pure stimulus act, can be any fraction of a total response pattern. Partly as a result of the organism's capacity to discriminate, the mediating response will tend to become as fractionated as possible without destroying its cue function.

> The greater the discriminatory capacity of an organism, the more reduced and implicit can become the 'detachable' reactions finally included in the stable mediation process. The higher the organism in the evolutionary scale, the finer the discriminations it can usually make and the less gross its representing processes. Similarly, the more mature and intelligent the human individual, the less overt his symbolic processes. The hosts of fine discriminations that characterize language behavior are Nature's farthest step in this direction (21, p. 398).

Of all possible human mediators, probably the most significant are verbal mediators. Research by Cofer and Foley and associates (5, 6, 7, 12), Riess (28, 29), and others demonstrates the significance for language behavior of mediating responses. Other investigations by Reed (23, 24, 25), Kendler and Karasik (15), and by Goss and associates (2, 11) show the importance of verbal mediators in concept formation. Studies by Eisman (9), Rhine (26, 27), and Staats and Staats (31) show how verbal mediators enter into attitudes.

The general model for a simple, first-order concept based on either verbal or nonverbal mediators is illustrated in Part A of Fig. 1. Many stimuli are conditioned to a common mediating response. By virtue of the common mediator, these stimuli become organized into one system. It is this system of associations between the stimuli and the common mediating process which is the concept. For the concept *gray,* the stimuli of Part A of Fig. 1 would be various acromatic shades all of which arouse the same mediating response. This response provides a stimulus generally conditioned in adults to the overt, verbal label *gray*.

A more complex or second-order concept is illustrated in Part B of Fig. 1. Each set of stimuli with their associated mediators (r_1's) are first-order concepts. In the second-order concept, the stimuli produced by the first-order mediators are themselves associated with another common mediator. One set of first-



A. FIRST–ORDER CONCEPT

B. SECOND–ORDER CONCEPT

C. ATTITUDE

FIG. 1. A stimulus-response explanation of concepts and attitude. S_e = evaluative stimuli; r_1 = a first-order mediator; r_{1_e} = a first-order evaluative mediator; r_2 = a second-order mediator; s = the stimulus produced by a mediator.

order stimuli in Part B of Fig. 1 might be different skin shades leading to the mediator representing *dark skin;* the second set of first-order stimuli might be various lip thicknesses which are associated with the mediator for *thick lips;* and each of the stimuli from these mediating responses may be associated with the mediator for *Negro* (r_2). This illustrates a second-order concept, but there is no attitude as yet because no evaluation is involved.

There is an attitude when the mediator of at least one of the first-order concepts is an evaluative reaction. Thus, attitudes are equivalent to that special class of concepts which is distinguished by the inclusion of an evaluative dimension. Part C of Fig. 1 is a simplified illustration of an attitude. Suppose, as in the example of a second-order concept, that one r_1 of this figure is the mediator for *thick lips,* that the second r_1 represents *dark skin,* and that r_2 is the mediator for *Negro*. Suppose also that the stimuli (S_e) of the evaluative dimension are *dirty, stupid,* and *rude* and that r_{1_e}, the evaluative reaction, is *bad.* Now the stimuli associated with Negro are *dark skin, thick lips,* and *bad.* The inclusion of an evaluative dimension identifies Part C of Fig. 1 as an attitude.

The mediating response produces a stimulus which may then become conditioned to other mediating or instrumental reactions. Part A of Fig. 2 shows some of the major associations possible between the attitude and these other mediating responses. It indicates how instrumental responses may depend upon a chain of associations involving two major types of mediators, namely, affective and verbal. Part A of Fig. 2 is not meant to imply that in any given instance all the illustrated associations between the attitude and verbal or affec-

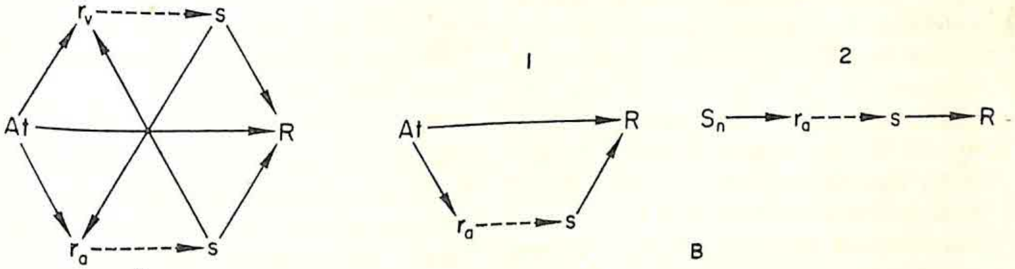


FIG. 2. A: Major associations between the attitude (At = the attitude as shown in Part A of Fig. 1), and verbal (r_v) and affective (r_a) mediators. B: A stimulus-response explanation of prejudice through displaced aggression.

tive mediators will be formed. Instead, it indicates that one or more of the associations may be learned by a given individual. For instance, as is shown in Part B-1 of Fig. 2, only some of the relations between the attitude, affect, and the instrumental response may be learned in an individual case.

One implication of Fig. 2 is particularly interesting. A reaction conditioned to the attitude may be aroused by a stimulus which is associated with affect but not with the attitude. This situation suggests an explanation for prejudiced responses through displaced aggression. Suppose an individual has a negative attitude towards Negroes. Suppose further that the affect of Part B of Fig. 2 is the mediator for *anger* and the instrumental response is aggression toward Negroes. Assume also, as Part B-1 of Fig. 2 indicates, that the attitude leads to the simultaneous arousal of anger and of aggression against Negroes. Eventually, through conditioning, anger alone can lead to aggression. Now, as indicated by Part B-2 of Fig. 2, any stimulus (S_n) which arouses anger will also have a tendency to evoke aggression towards Negroes, depending upon the degree to which this stimulus is supported by other stimuli. As a result, there is a tendency to displace aggression from S_n, the real source of anger, to Negroes.

Other specific predictions follow from the general relations shown in Part A of Fig. 2. The relationship between verbal mediating responses and affect suggests one reason why attitudes are commonly accompanied by affective reactions. If a given verbal mediator, through past experience, has become associated with affect, then the arousal of this mediator by an attitude will also lead to an affective response: the attitude leads to the verbal mediator and the verbal mediator is associated with affect. Suppose, for instance, that an attitude supplies the conditioned stimulus which evokes the verbal mediator for *green vomit*. Then, insofar as this mediator evokes an emotional reaction, the attitude will be accompanied by affect. This particular verbal mediator was used as an example because in some cases it will make the illustration more vivid by actually arousing an emotional reaction in the reader.

There is probably a more important reason why attitudes are commonly linked to affect. It is quite often true that the experiences through which the attitude is developed occur in a charged atmosphere. In this case, the mediating response of the attitude is continuously aroused in conjunction with an emotional reaction, and the stimulus produced by the mediating response becomes a conditioned stimulus capable of evoking affect. As a result, the expression of this attitude is accompanied by an emotional response.

A METHOD FOR STUDYING ATTITUDE DEVELOPMENT

One of the main values of the concept-formation approach to attitude is that it suggests a simple laboratory method for studying attitude development. If an attitude is a kind of concept, then its acquisition may be investigated by the methods of concept formation. Such a method, depending on an evaluative dimension in the concept, has been employed by the writer. In order to construct an evaluative dimension, 100 *S*s were asked to rate a large number of descriptive traits in terms of their desirability in a person. Some example traits are: ambitious, boring, calm, intelligent, obedient, and vain. All traits showing less than 90% agreement as to their desirability were discarded. The remaining traits can be employed to study the development of an attitude and to measure its strength and resistance to change. To do this, a group of *S*s is told that anthropologists, missionaries and physicians have visited a group of people about which little is known, and that all the visitors agree on certain characteristics of this group of people. When these instructions are concluded, *S*s are shown a number of traits, one by one, and are asked to predict whether each trait is or is not characteristic of the people. After each prediction they are told whether or not the word is indeed characteristic. For example, to develop a positive attitude, *E* would say "characteristic" after presenting each trait previously rated as desirable and "not characteristic" for words previously rated undesirable.

If *S* has learned the positive attitude, which might be labeled "pro-ness," he would be expected to respond accordingly. Hence, when he has completed the first list of traits, *S* is asked to continue responding in the previous manner to a second list of words without being told which traits are characteristic or not characteristic. Responses under this condition are taken as a measure of attitude strength. It is assumed that a strongly acquired, positive attitude will lead to a high proportion of positively rated traits being called "characteristic" and to most negatively rated traits being called "not characteristic."

Finally, resistance to attitude change is measured by using a third list of traits with *E* again saying "characteristic" or "not characteristic" after each prediction. But this time *E*'s responses are reversed. For example, if *S* has previously acquired a positive attitude, change may be accomplished by calling "not characteristic" all positive traits in the third list and "characteristic" all negative traits.

This method is derived from the concept-formation approach to attitudes. The meaning of the many traits seen in a single session varies considerably, as can be observed from the examples given above. Although the traits are not synonyms, they do exhibit one common aspect, the evaluative element. The *S*s are able to respond correctly when this evaluative element is consistently reinforced by *E* because the traits are stimuli which are capable of arousing a common, evaluative mediating response. The mediating process, in turn, determines *S*'s further predictions of "characteristic" or "not characteristic."

The stimulus function of mediating responses in attitude development is demonstrated by two experiments which also illustrate the mediating response approach as applied through the concept-attitude method. A negative attitude was developed in both investigations. In one study (27), consistency of reinforcement was defined as the sum of the number of positive traits called "not characteristic" by *E* and the number of negative traits called "characteristic," divided by the total number of traits used during attitude development. Using five levels of consistency, it was shown that both attitude development and strength increase as a function of consistency of reinforcement, as would be predicted from the approach to attitude development based on mediating responses. The results for attitude change also suggest that attitudes exhibit properties of mediating responses. Research on acquired distinctiveness of cues (18) indicates that strong habits reverse more easily than weak ones when the cues arousing them have become distinct through learning. The distinctiveness of the stimulus dimension is believed to be related to the degree to which the stimuli are conditioned to a mediating process. The evaluative element of the stimulus traits should be most distinct when reinforcement is fully consistent; consequently, attitude change should be easiest under this condition. The *S*s receiving the greatest consistency of reinforcement did in fact show least resistance to change.

A second experiment (26) illustrating the concept-attitude method investigated the effect on attitude development of peer responses. The traits were used with 100% and 84% consistency of reinforcement to develop a negative attitude. Some *S*s made their predictions after first hearing the responses of three fellow students who were *E*'s confederates, and other *S*s made their predictions in private. The confederates' responses prior to *S*'s predictions were always or almost always in contradiction with the statement *E* made after each prediction. Thus, when the confederates said a trait was characteristic, it was pretty certain that *E* would say the trait was not characteristic. The *S*s who heard the confederates' responses learned the attitude with greater ease than those who responded in private. Apparently, peer responses provided cues which were an aid to *S* in attaining the attitude. This seemed to occur mainly with high consistency of reinforcement. With less than full consistency of reinforcement, peer responses were predictive of *E*'s behavior for a majority of the traits, but not all of them. When this ambiguity is introduced, cues from peer responses do not affect attitude development at a statistically significant level, although there is a slight trend in this direction.

These results may be explained by considering the effect of peer responses upon the arousal of the common, evaluative mediating response. With 100% consistency of reinforcement, a negative mediating reaction is aroused by a trait just prior to reinforcement and in conjunction with peer responses of "characteristic." Similarly, a positive mediating reaction is aroused in conjunction with peer responses of "not characteristic." In time, as a result of conditioning, the peer responses can be expected to help cue the appropriate positive or negative mediating response. With 84% consistency of reinforcement, the trend of conditioning is the same, but a minority of trials tend to establish competing, mediating responses: the same peer reactions are sometimes associated with positive mediating responses and sometimes with negative. Under these conditions, peer responses are a less stable and less predictive aspect of the stimulus environment. This would account for the statistically significant effect from peer responses with high consistency of reinforcement, which is diminished with lesser consistency.

Staats and Staats (31) have also studied attitudes using an approach similar to that employed in the preceding two studies. The associated given names, Tom or Bill, and nationalities, Dutch or Swedish, with positive or negative evaluative words. Subsequent ratings indicate that the given names and nationalities took on the same evaluative flavor as the words with which they were associated. The researchers explain their results by pointing to the common, evaluative mediating response made to the words which are paired with the nationalities and names. It is assumed that through conditioning the names (or nationalities) would also lead to the mediating response aroused by the words with which they were paired. This was reflected when the names (or nationalities) were rated on the evaluative scale.

The studies derived from the concept theory of attitudes illustrate the possibility of bringing attitude development under laboratory scrutiny. Despite the generally recognized significance of attitude development, as distinguished from attitude change, the literature reveals relatively little laboratory evidence bearing upon its nature. It is hoped that the concept theory of attitudes and the method it implies will help provide a means through which attitude development will be more accessible to experimental investigations.

SUMMARY

A theory of attitude development based upon concept formation was discussed, and similarities between this view and earlier meanings of attitude were pointed out. An attitude was defined as a concept with an evaluative dimension, and the mediating response was employed to explain attitude learning. A method for inducing an attitude in the laboratory was described along with illustrative studies employing this method.

REFERENCES

1. CANTRIL, H. Attitudes in the making. *Understanding the child,* 1934, **4**, 13–14, 18.
2. CAREY, J. E., & GOSS, A. E. The role of verbal labeling in the conceptual sorting behavior of children. *J. genet. Psychol.,* 1957, **90**, 69–74.
3. CHEIN, I. Behavior theory and the behavior of attitudes: Some critical comments. *Psychol. Rev.,* 1948, **55**, 175–188.
4. CLARK, K. B., & CLARK, MAMIE P. Racial identification and preference in Negro children. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1952.
5. COFER, C. N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.,* 1942, **49**, 513–540.
6. COFER, C. N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.,* 1943, **32**, 168–175.
7. COFER, C. N., JANIS, MARJORIE E., & ROWELL, MARY M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.,* 1943, **32**, 266–269.
8. DOOB, L. W. The behavior of attitudes. *Psychol. Rev.,* 1947, **54**, 135–156.
9. EISMAN, BERNICE S. Attitude formation: The development of a color preference response through mediated generalization. *J. abnorm. soc. Psychol.,* 1955, **50**, 321–326.
10. FARIS, E. The concept of social attitudes. In K. Young (Ed.), *Social attitudes.* New York: Holt, 1931.
11. FENN, J. D., & GOSS, A. E. The role of mediating verbal responses in the conceptual sorting behavior of normals and schizophrenics. *J. genet. Psychol.,* 1957, **90**, 59–67.
12. FOLEY, J. P., JR., & MATHEWS, M. A. Mediated generalization and the interpretation of verbal behavior: IV. An experimental study of the development of interlinguistic synonym gradients. *J. exp. Psychol.,* 1943, **33**, 188–200.
13. HOROWITZ, E. L. The development of attitude toward the Negro. *Arch. Psychol.,* 1936, No. 194.
14. HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.,* 1920, **28**, No. 1 (Whole No. 123).
15. KENDLER, H. H., & KARASIK, A. D. Concept-formation as a function of competition between response produced cues. *J. exp. Psychol.,* 1958, **55**, 278–283.
16. KRECH, D. Attitudes and learning: A methodological note. *Psychol. Rev.,* 1946, **53**, 290–293.
17. KRUEGER, E. T., & RECKLESS, W. C. *Social psychology.* New York: Longmans, Green, 1931.
18. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.,* 1950, **40**, 175–188.
19. LEEPER, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
20. NELSON, E. Attitudes: I. Their nature and development. *J. genet. Psychol.,* 1939, **21**, 367–399.
21. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
22. OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.,* 1955, **62**, 42–55.
23. REED, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.,* 1946, **36**, 71–87.
24. REED, H. B. The learning and retention of concepts: II. The influence of length

of series. III. The origin of concepts. *J. exp. Psychol.,* 1946, **36**, 166–179.

25. REED, H. B. The learning and retention of concepts: IV. The influence of complexity of the stimuli. *J. exp. Psychol.,* 1946, **36**, 252–261.

26. RHINE, R. J. The effect of peer group influence upon concept-attitude development and change. *J. soc. Psychol.,* in press.

27. RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept attitude as a function of consistency of reinforcement. *J. exp. Psychol.,* 1958, **55**, 524–529.

28. RIESS, B. F. Semantic conditioning involving the GSR. *J. exp. Psychol.,* 1940, **26**, 238–240.

29. RIESS, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.,* 1946, **36**, 143–152.

30. SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.,* 1932, **42**, No. 4 (Whole No. 191).

31. STAATS, A. W., & STAATS, CAROLYN K. Attitudes established by classical conditioning. *J. abnorm. soc. Psychol.,* 1958, **57**, 37–40.

(Received January 28, 1957)

Psychological Review
Vol. 65, No. 6, 1958

SYMPTOMS AND SYMPTOM SUBSTITUTION

AUBREY J. YATES

University of New England, Australia

In recent years there has been a number of attempts to derive rational methods of symptomatic treatment of the behavioral disorders, based mainly on learning theory (10, 12, 13, 15, 18, 19) and information theory (2). Nevertheless, symptomatic treatment has not been favorably considered by most clinical psychologists, and the failure to appreciate its possibilities is perhaps best shown by the almost total neglect of Dunlap's classic work (3) on habit-formation and the remedial treatment of undesirable habits. The object of this paper is to point out some misconceptions on which objections to symptomatic treatment are based, and to show that such treatment can in fact be firmly founded on modern learning theory.

The argument that symptomatic treatment is a waste of time has had as its strongest proponent the psychoanalytic school, which has insisted that symptoms are but the surface indicators of underlying conflicts, anxieties, etc. This argument has been accepted by most "dynamic" psychologists, who concluded that removing the symptom without also effectively treating the anxiety or conflict underlying it would lead to symptom substitution. For example, even if tics could be treated directly and extinguished, the patient would quickly develop either new tics or some other form of symptomatic response. Even orthodox psychiatrists who have little patience with Freudian psychodynamics have accepted this point of view. Thus Kanner writes that "symptomatic therapy which does not care what the complaint is a 'symptom' of is rarely successful and leads to unwarranted pessimism" (9, p. 244). Mowrer has summed up the position neatly and abusively:

> . . . the behavioral manifestations which we now call "symptoms" are not the essence of neurosis and . . . the modes of treatment that are aimed at their direct alleviation are as ill-considered theoretically as they are futile practically (14, p. 620).

This point of view has become so widely accepted that until recently almost no symptomatic treatment was carried out, except in the educational field of remedial teaching and speech therapy. However, the distinction between a "fundamental" underlying anxiety and a "superficial" surface symptom is open to a number of objections.

1. The term "symptom" has been employed very loosely in psychology. In general medicine a clear distinction is made "between symptoms, that is disorders described by the patient, and physical signs or the deviations from the normal revealed by examination" (7, p. 1). In other words, a symptom is a subjectively experienced abnormality, a sign is an objective indication of abnormality. In many cases the symptom may, of course, also be a sign.

Mowrer, having accepted, as we have seen, Freud's conclusion with regard to the meaning and function of symptoms, was considerably embarrassed by his own empirical finding that "symptomatic" treatment of enuresis was not only 100% successful with regard to the symptoms, but was not followed by symptom substitution in a single case! Instead, however, of rejecting Freud's hypothesis, Mowrer took the extraordinary step of denying that enuresis was a symptom in any of the cases he treated. He distinguished between habits which were also symptoms, and habits which were not. Thus,

> "symptoms" differ from ordinary "habits" primarily in that the former are motivated by and are perpetuated because they lessen the drive of anxiety, whereas "normal" habits are motivated by and are perpetuated because they lessen other drives, such as hunger, thirst, cold, fatigue, sex, fear, etc. (14, p. 546).

Now the above distinction is clearly a meaningful and possibly a valid one. As far as the author is aware, however, Mowrer has not presented any evidence that the distinction can be sustained in the form in which he has put it, and his own experimental results seem strongly to contraindicate the validity of any such distinction. Mowrer's acrobatics in this regard are as good an example as any of an a priori clinical bias being allowed to outweigh cogent experimental evidence.[1]

It is clear that, as used by Freud (6), Mowrer (14), and other "dynamic" psychologists, the symptom is regarded as a response to some hidden, basic conflict or anxiety. It will be suggested that this "vertical," dualistic approach in unnecessary and should be replaced by a "horizontal," monistic approach, as described below.

2. The "dynamic" approach ignores the distinction between neuroticism and neurosis (4). Neuroticism may be conceptualized as an innate predisposition to develop a neurosis under certain specifiable conditions. This conceptualization is at least as plausible as the alternative outlined above. But, if such a predisposition exists, then it follows that a neurosis *is* symyptoms and nothing else, i.e., that a particular neurosis consists of a particular set of learned responses; and that treatment on psychological grounds can only be symptomatic, since treatment of the predisposition must ultimately be by genetic or chemical means.

3. Finally, it should be noted that belief in symptom substitution seems to be based largely on clinical experience. Considering the significant role such a distinction has played in clinical psychology, experimental demonstration of its existence is singularly lacking. We conclude, therefore, that the reasons generally given for rejecting symptomatic treatment in favor of "dynamic" treatment are not soundly based. Nevertheless, the definition of a neurosis as a system of learned responses is not in itself incompatible with the notion of symptom substitution. We must now turn to the question whether there are any theoretical reasons to suppose that symptom substitution would take place, if symptomatic treatment were to be undertaken. It will be maintained that learning theory suggests both that symptomatic treatment is a rational procedure and that symptom substitution would not be expected except under certain specifiable conditions.

This may perhaps be most easily shown by reference to some recent work on the symptomatic treatment of tics (cf. 19 for details). A model was constructed to account for the genesis of some tics in terms of two-factor learning theory. It was hypothesized that the tic is a conditioned avoidance response to anxiety, the latter being also a learned response (the conditioned form of the original fear response). A method of treatment was derived in terms of the Hullian concepts of I_R and $_sI_R$, making use of massed practice.[2] With regard to this theory, the following points are specifically relevant to this discussion:

---

[1] Mowrer admits that it is impossible to "know from the incontinence alone that it is a symptom rather than simply a reflection of pedagogical inadequacy" (14, p. 414). Presumably, therefore, decision as to whether the incontinence is a symptom or a habit can be determined only by carrying out "symptomatic" treatment and seeing what happens.

[2] In this study (19) it was found convenient to use the concepts of I_R and $_sI_R$ in deriving a method of treatment. This aspect could also, however, have been conceptualized in terms of two-factor learning theory, as Mowrer has shown in connection with studies in the field of general motor behavior (14, p. 169, Footnote 4).

a. Both the anxiety and the tic are learned responses. Now it would be a logical error to suppose that the fact that the tic is a response *to* the anxiety necessarily means that the anxiety must therefore be more fundamental or that it must receive prior treatment. It seems more reasonable to regard the situation simply as one in which two sets of responses (anxiety and the response to anxiety) are learned over time. This may be termed the "horizontal" approach, as opposed to the dualistic "vertical" approach rejected above.

b. From this point of view it does not much matter which aspect of behavior (anxiety or tic) is treated first. The fact is, however, that in our present state of knowledge it is easier to treat the tic than the anxiety. The former is more readily manipulated and measured, and, since it usually involves motor behavior, is easily dealt with in relation to the concepts of learning theory. It turns out, however, that in treating the symptom we are also treating the anxiety. For suppose the symptom is destroyed; then the anxiety, instead of being reduced by it, will persist. But since in fact no traumatic event occurs if the anxiety is allowed to persist, the latter, being itself a learned response, will begin to extinguish. This is what in fact seems to happen, though the principle of partial irreversibility of traumatic anxiety (17), if valid, suggests that extinction of the anxiety would never be complete.[3]

This formulation is, of course, derived directly from Mowrer's two-factor theory of learning. That removal of the symptom will affect the learned anxiety is hinted at by Mowrer, when he explains why a symptom

> is at one and the same time both self-perpetuating and self-defeating. It is self-perpetuating because it is reinforced by the satisfaction provided through the resultant anxiety reduction; and it is self-destructive in that *it prevents the individual from experiencing the full force of his anxiety and being modified by it in such a direction as to eliminate the occasion for the anxiety* (14, p. 535, Footnote 3; italics not in original).

Supposing, therefore, both the symptom and the anxiety to be alleviated, we would be left with a basically neurotic person without a current behavior disorder. But, in terms of learning theory, there would be no reason to suppose that this person would develop new anxieties or symptoms unless put into a new traumatic situation. This would, of course, involve knowledge of what situations would be likely to be traumatic for a particular patient.[4]

Such empirical evidence as there is supports the position outlined above. Symptom substitution has not been found in those instances where a single patient has been subjected to intensive investigation and symptomatic treatment (e.g., 8, 12, 19); nor in those few instances where large-scale follow-up has been carried out on groups of subjects. For example, two recent investigations (1, 11) of the results of symptomatic (*pace* Mowrer) treatment of enuresis indicate 70%–80% success, and a failure to find symptom substitution over periods up to two years. Mowrer himself (13) reported 100% success. The critical point seems to be that we are faced in such cases with a vicious circle (as Mowrer has indeed pointed out): the tic, for example, is a response to anxiety, but performance of the tic increases the general level of drive (irritability, annoyance, etc.), which increases the rate of responding, and so on. The important factor seems to be the breaking of this vicious circle; it does not much matter whether this is accomplished by treating the anxiety or the symptom, since both react upon each other. Thus, in the shadowing method described by Cherry and Sayers (2) for the treatment of stammering, the mere fact of demonstrating to the patient that he can speak normally (even though under special conditions and before treatment has started) usually has a very striking effect on the patient's general well-being. In many instances it seems that the symptom itself generates a good deal of anxiety, which can be alleviated by a direct attack on the symptom.

---

[3] The nature of traumatic avoidance learning is a very complex problem, which has been greatly oversimplified here. The reader is referred to recent papers by Solomon and Brush (16) and by Solomon and Wynne (17) for comprehensive discussions of the current position.

[4] In the case of the patient treated by the author, for example, any situation involving anaesthesia (such as dental or medical operations) would be potentially traumatic.

The seriousness of the present situation is clearly demonstrated by Eysenck's recent contention (5, p. 271) that, in spite of the experimental validation of symptomatic treatment of enuresis, this technique is not used in any child guidance clinic in Great Britain. It is obvious that a reorientation is highly desirable, in which the "vertical" approach is abandoned, the use of the term "symptom" dropped altogether, and the patient considered as an individual who has developed a series of learned responses to certain situations. Clinical psychologists may then perhaps begin to apply that large body of knowledge and theory to be found in general psychology in which they are specially trained, and apply the principles of behavior to the abnormal field as well as to the normal.

REFERENCES

1. Baller, W., & Schalock, H. Conditioned response treatment of enuresis. *Except. Child,* 1956, **22**, 233–236.
2. Cherry, C., & Sayers, B. McA. Experiments upon the total inhibition of stammering by external control and some clinical results. *J. psychosom. Res.,* 1956, **1**, 233–246.
3. Dunlap, K. *Habits. Their making and unmaking.* New York: Liveright, 1932.
4. Eysenck, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.
5. Eysenck, H. J. *The dynamics of anxiety and hysteria.* London: Routledge & Kegan Paul, 1957.
6. Freud, S. *The problem of anxiety.* New York: Norton, 1936.
7. Holmes, G. *Introduction to clinical neurology.* Edinburgh: Livingstone, 1946.
8. Jones, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.,* 1956, **52**, 414–419.
9. Kanner, L. *Child psychiatry.* Springfield, Ill.: Charles C Thomas, 1935.
10. Liversedge, L. A., & Sylvester, J. D. Conditioning techniques in the treatment of writer's cramp. *Lancet,* 1955, **2**, 1147–1149.
11. Martin, B., & Kubly, D. Results of treatment of enuresis by a conditioned response method. *J. consult. Psychol.,* 1955, **19**, 71–73.
12. Meyer, V. The treatment of two phobic patients on the basis of learning principles. *J. abnorm. soc. Psychol.,* 1957, **55**, 261–266.
13. Mowrer, O. H. Enuresis: A method for its study and treatment. *Amer. J. Orthopsychiat.,* 1938, **8**, 436–459.
14. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald Press, 1950.
15. Raymond, M. J. Case of fetishism treated by aversion therapy. *Brit. med. J.,* 1956, **2**, 854–857.
16. Solomon, R. L., & Brush, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Univer. Nebraska Press, 1956. Pp. 212–305.
17. Solomon, R. L., & Wynne, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychol. Rev.,* 1954, **61**, 353–385.
18. Wolpe, J. Experimental neuroses as learned behavior. *Brit. J. Psychol.,* 1952, **43**, 243–268.
19. Yates, A. J. The application of learning theory to the treatment of tics. *J. abnorm. soc. Psychol.,* 1958, **56**, 175–182.

(Received January 30, 1958)

Psychological Review
Vol. 65, No. 6, 1958

TRANSFER OF TRAINING AND ITS RELATION TO PERCEPTUAL LEARNING AND RECOGNITION[1]

JAMES M. VANDERPLAS

Washington University

It is the purpose of this paper to bring together several current concepts arising from the study of transfer of training and to consider their implications for the study of perceptual learning and recognition. A number of studies have indicated an increasing emphasis on relations between these two areas (1, 2, 3, 12, 13, 21, 22, 27, 38), and several theoretical formulations have been advanced to account for certain phenomena in perceptual learning and recognition in terms of existing principles of transfer (12, 21, 26, 37). In several instances, however, the reverse of this latter kind of account has been advanced; that is, certain phenomena in transfer of training have been held to support theoretical formulations grounded upon assumptions involving processes of perceptual learning (3, 15, 16, 17, 20).

In the first kind of view, recognition or discrimination tasks have been employed to provide evidence for hypotheses concerning changes in stimulus-response relations which take place during learning. This view is illustrated nicely by the study by Arnoult (2). Four groups of subjects learned to associate verbal labels with nonsense forms. In one group the labels were nonsense syllables; in a second group the labels were girls' names; in a third group the labels were descriptive names given the forms by an independent group; in the fourth group the labels were names invented by the subjects themselves. Following this labeling practice, the subjects were given a recognition test, in which they were required to select from a group of forms those for which they had learned the labels. Recognition was found to increase as a function of the degree to which the responses in the paired-associates task (learning labels) reflected associations made to the forms. The implication from this kind of study is that recognition, as a perceptual task, is dependent (at least in part) upon the degree to which responses, in this case labeling responses, transfer from the first task to the task of recognition. Clearly, then, from this point of view, the process of recognition is intimately related to what may be regarded as stimulus-response relations, consistent with long-standing associationistic theories.

In the second kind of view, the converse of the above statement is usually made; transfer comes about as a function of underlying perceptual processes. This view finds support in studies of the kind made by Gagné and Baker (15). In this study, subjects were given preliminary practice in associating a different letter to each of four different lights. Following this practice the subjects engaged in a second task in which the same lights were used as stimuli as were used in the first task, but the responses were changed to require the subjects to manipulate a different switch when each of the four lights appeared. The scores (time and errors) were found to vary in a way which was related to the amount of practice in the first task. The rate at which the subjects learned to respond by manipulating the correct switch increased as a function of practice in associating labels, and significant differences in errors were found as a function of the amount of first-task practice. These results were interpreted by the investigators as supporting the view that practice in labeling brought about a reduction in generalization in responses among the various intratask stimuli. In their view, the practice increased the degree to which the stimuli were differentiated from each other. A similar point of view, although based on a somewhat different mechanism, is contained in the well known hypothesis of Miller and Dollard (33) of acquired distinctiveness of cues.

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 41 (657)—47, monitored by the Operator Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

Although these two kinds of views have not always been stated explicitly, they appear in a variety of theoretical formulations (to be examined later in this paper) and in a number of studies of transfer of training (9, 11, 29, 30), stimulus predifferentiation (1, 3, 6, 15, 16, 20, 38), acquired distinctiveness of cues (21, 33, 39), and in studies of perceptual learning (7, 12, 13, 38). It is evident that these views are closely related. The first view mentioned above implies a close relation between verbal and motor responses, *qua* responses, and responses which are presumed to reflect or to influence perceptual responses. The second view implies that verbal and motor responses may be related to stimulus variables which influence perceptual processes. Thus, we might expect that perceptual learning and recognition are activities which are understandable in terms of transfer of training and, conversely, that transfer of training is dependent upon (or, in the least, involves) processes similar to those encountered in studies of perceptual learning and recognition. However, as obvious as these relations, stated in this broad sense, appear at first, several difficulties arise almost immediately to plague the investigator who attempts to interpret in any clear fashion the results of experiments in these areas.

Taken as a group, the studies cited above tend to support a common conviction among psychologists that transfer is involved in practically all phenomena beyond those of the simplest reflex variety. Conversely, however, they also support an equally common conviction that perception is involved in practically all forms of activity, in one way or another.

But if one examines carefully the frameworks of these convictions, one encounters several difficulties. First of all, the principles of transfer of training themselves do not find unequivocal support in experiments designed to test them. For example, a frequently cited principle of transfer, made most explicit in the formulation of Osgood (36), is that when responses are different in two tasks, transfer will occur from one task to the other in a fashion inversely related to the degree of similarity of the responses in the tasks. Positive transfer will occur when the responses are similar, and zero or negative transfer will occur when the responses are unrelated or antagonistic. But a direct test of this principle, in an experiment by Bugelski and Cadwallader (10), in which responses were varied over a wide range, yielded data contradictory to the principle.

Second, specification of the ways in which responses, *qua* responses, in transfer formulations on the one hand, and responses presumed to reflect recognition on the other hand, are related, is far from clear. One not only finds perceptual responses interpreted as instrumental activity dependent upon inferred responses of an identification type (26), but one also finds instrumental responses interpreted as resulting from processes involving changes in perception (20). These differences in interpretation of the activities only lead to confounding of the investigator interested in a clear statement of principles.

Third, the variables which are employed to represent the various concepts in these views are often subject to interpretations which support alternative views with almost equal likelihood. In Arnoult's experiment, for example (2), the response of recognition (selection of the correct nonsense form) could be interpreted as resulting either from reduction in generalization or from facilitation of the correct response (*qua* response), brought about directly by way of the intervening identification response that had been made during the practice in labeling. A similar interpretation could be made of the results of the experiment by Gagné (15), cited above. The correct switching response could result either from reduction in generalization (predifferentiation), acquired distinctiveness (33), or from an increase in likelihood of the correct response due to its relation to the response in the first task (9, 30). Thus, although recognition may be interpreted broadly as resulting from one or another of the principles of transfer, the difficulties mentioned here suggest that it is in good order to reexamine some of the hypotheses generated from the various studies.

THE RELATION BETWEEN TRANSFER OF TRAINING AND RECOGNITION

Despite a considerable divergence in point of view and method between studies of transfer and recognition, they involve similar experimental operations.[2] Their similarities may be illustrated by considering the following paradigms.[3]

Recognition

Training or experience is provided by engaging the subject in a task in which he is asked to observe, label, draw, or otherwise respond to a form, a pattern, or other stimulus. Several such stimuli may be used. Following training, he may be asked in a second task to select those stimuli used in the first task from among a group containing both the old and some new stimuli.

This paradigm has been employed in a number of studies designed to test hypotheses about effects of predifferentiation or acquired distinctiveness training upon transfer, in both verbal and motor tasks (1, 2, 3, 9, 12). In a study by Vanderplas and Garvin,[4] for example, training was employed in which subjects were required to learn nonsense labels for a series of nonsense forms. Following this labeling practice, a test of recognition was employed in which the subjects were required to select the forms for which labels were learned when imbedded in a set of forms containing both the old forms and several similar ones. Several conditions were employed, including amount of practice and certain physical variations in the forms. The study, however, illustrates the use of the paradigm as a test of the hypotheses mentioned above. As another example, the study by Kurtz and Hovland is of interest: in their study (27), subjects who verbalized the names of common objects were found to recognize them with greater certainty in a second task than a group who did not so verbalize. The results thus supported the hypothesis that the naming response increased the probability of correct recognition. For the present purpose, the study illustrates once again the point of view, mentioned earlier, that recognition may be influenced by the relation that it bears to responses made in the initial task. The paradigm, as described above, has also been employed in studies of concept formation (8) and in the more classical experiments on form and pattern perception (40, Ch. 23). In studies where this paradigm has been employed, it is usually inferred that the response of recognition is dependent upon its relation to responses (most often of an identification type) made during practice on an initial task, and thus closely related to existing principles of transfer of training.

---

[2] Difference in point of view concerning similar operations may be illustrated by considering a situation in which a person encounters a woman, approaches her, and says, "Hello, Mary." If the woman were not in fact named Mary, the response might be regarded as a false recognition. But if the woman were one of a group of women and if the person had the task of learning the names of each woman, the response "Hello, Mary" might be regarded as a case of positive transfer, provided that another woman (in a previous situation) had been correctly called Mary by the person.

[3] The experimental controls necessary in this and the following paradigm are not cited here, since they are intended to show similarities. Controls vary with the phenomena under investigation, and serve to distinguish experiments rather than points of view.

[4] A report of this research, entitled "Complexity, Association Value, and Practice as Factors in Recognition Following Paired-Associates Verbal Training," is now in preparation.

Transfer of Training

Training is provided in which the subject is required to make a verbal or motor response to a set of stimuli, such as lights, tones, words, nonsense syllables, or even forms or patterns. Following this training he is required to perform a second task which may be of one or more of the following types (symbolized in S–R terms, for convenience):

> *Type SRSR:* The subject is asked to respond in the second task as he did in the first, to the same stimuli.
> *Type SRsR:* The subject is asked to respond in the second task as he did in the first, but to stimuli different from those of the first task.
> *Type SRSr:* The subject is asked to respond differently in the second task to the same stimuli used in the first.
>
> *Type SRsr:* The subject is asked to respond differently in the second task to stimuli different from those used in the first.

In these paradigms, instructions to the subject may include directions to respond as above, or he may be given only general instructions such as those provided in paired-associates training where he is asked to attempt to anticipate each response item in a list.

The paradigms have been extensively employed in studies of predifferentiation, acquired distinctiveness of cues, transfer, and retention. It is often inferred from the experimental results, in studies employing these paradigms, that the response employed in the first task facilitates performance of the second if transfer is demonstrated. For example, in a study by McAllister (30), it was shown that when a verbal response employed in a first task had some degree of "relevance" to a corresponding motor response in a second task, transfer increased. The relevance of the first-task response to the motor response was by way of the denotative character of the verbal response (e.g., "up-left," and the corresponding movement of a lever in the star discrimeter).

When these paradigms are thus considered, in terms of the stimulus-response relations in the two tasks, several points of close correspondence in the points of view mentioned earlier become apparent. First of all, in the recognition paradigm, the subject is asked to respond in the second task (usually by implication) in two ways: (*a*) as he did in the first to the same stimuli, and (*b*) differently to different stimuli. If he responds correctly with the responses that he learned during first-task training, it is predicted by the transfer theorist that the response of recognition will be facilitated. Conversely, if recognition is found to improve following practice in labeling, say, it is inferred that the response of labeling facilitated recognition. These requirements, implicitly made in the recognition paradigm, are tantamount to imposing, simultaneously, tasks of Type *SRSR* and *SRsr* of the transfer paradigm. Thus, experiments on recognition which implicitly impose these requirements may be regarded as forms of transfer tasks which involve simply differential response. Conversely, transfer experiments which impose similar tasks may support a view that differential responses arise from processes of recognition.

To emphasize the relations between these paradigms, consider also the following.

1. Tasks of Type *SRSR* require retention of a stimulus, a response, or a specific S–R relation from the first task to the second. They also require recognition of the stimulus in order that the correct response can be made.

2. Tasks of Type *SRsR* may require generalization of a response from first-task stimuli to similar stimuli in the second. Failure of recognition or differential discrimination would result in positive transfer.

3. Tasks of Type *SRSr* may require a different response to the same stimuli in the second task. Failure to recognize a stimulus as old could result in positive transfer; i.e., the subject would respond differently, as required. Correct recognition would result in negative transfer, since the subject would respond with the old response, and interference would result.

4. Tasks of Type *SRsr* require differential responses to different stimuli in the two tasks. Recognition should result in just such performance. Zero transfer would be predicted, however, since no old stimuli are present to evoke the recognition response.

To summarize, recognition and transfer may be regarded as directly related, inversely related, or not related at all, depending upon the relations between the stimuli, responses, or both, in the two tasks. In general (but only in general), they may be regarded as related by way of the common requirement of differential response to different stimuli and its relation to activities which reflect recognition. With these points in view, let us examine some current principles of transfer as they relate to the study of recognition.

APPLICABILITY OF PRINCIPLES OF TRANSFER TO THE STUDY OF RECOGNITION

A number of investigators have attempted to account for transfer of training in verbal and motor tasks by application of hypotheses derived from studies of perceptual phenomena, primarily those involving discrimination. At the same time, hypotheses arising out of association theory (31, 32, 37), studies of concept formation (8), and S–R formulations (21) have been invoked to account for data otherwise presumed to reflect perceptual processes. Although they have seldom been advanced in pure form, we may distinguish three divergent points of view on these questions.

First, classical theory regards transfer to be a function of S–R relations, and to be dependent upon similarity relations between stimuli or responses in the two tasks. This view is clearly seen in the transfer and retroaction model of Osgood (36), to which we have already made reference. A second view is seen in the studies of stimulus predifferentiation. Transfer is now regarded as a function of changes in discriminability of stimuli, either as a result of reduction of generalization gradients (18) or as a result of increased distinctiveness arising from the addition of response-produced cues (21, 33). This view is presented in the well known hypothesis of Miller and Dollard (33), and in the extended formulation of Goss (21). A third view regards transfer as resulting from certain mediational processes, including verbal associations to task stimuli. This view has received emphasis in the recent discussion by Postman, who holds to the point that perceptual learning may be encompassed by association theory, by way of an account which includes verbal response processes (37). To what extent each of these divergent views is relevant to the study of recognition is a question of some concern. We shall next attempt to examine these formulations with this question in mind.

S–R SIMILARITY RELATIONS AND TRANSFER

Osgood, in his treatment of transfer (36), regards three principles to be of primary importance in determining the direction and extent of transfer: (*a*) when responses in the two tasks are (functionally) identical, positive transfer is obtained and increases with increasing similarity of the stimuli; (*b*) when the stimuli in the two tasks are (functionally) identical, with responses varied, negative transfer is obtained and decreases with increasing similarity of the responses; (*c*) when both stimuli and responses are varied, negative transfer is obtained and increases with increasing stimulus similarity.

It is apparent from the foregoing discussion of the paradigms cited that these principles, paraphrased from Osgood's discussion of the model (36, pp. 134–135), are partly principles of recognition. However, some qualifications need to be mentioned. First, recognition may be regarded as a case of transfer in terms of the paradigm for tasks of Type *SRSR*, since both stimuli and responses are identical and should lead to correct recognition. But failure of recognition could also be regarded as positive transfer, since the identical response, made to a new stimulus, would lead to an incorrect selection (recognition) response. The second principle would apply to recognition, if a subject first learned a label for, say, a form and then was asked to press a switch when that form again appeared. This would correspond to a task of Type *SRSr*, and transfer would depend on the similarity between the label and the switch-pressing response. But a question might arise over one's definition of similarity between the responses. A number of experiments, for example, have indicated that positive transfer may occur in cases where the stimuli of the two tasks are identical but where the responses are clearly different (6, 9, 11, 14, 29, 30).

A third caution in the interpretation of these principles involves the implicit assumption that transfer is related to discrimination. In terms of the model, stimuli which are least discriminable are most likely to evoke responses which may be interpreted as denoting positive transfer, but also as reflecting failure of recognition. This point need not be of major concern, but care needs to be exercised in any broad generalizations about the applicability of these kinds of principles.

Thus, we see that S–R formulations may be used in part to account for some of the findings from studies of recognition. On the other hand, some facts of recognition tasks are not in accord with these principles—indeed, even some facts of transfer are not, as noted earlier (10)—and additional hypotheses may be needed. Some of these added hypotheses may be found in studies of transfer as related to discrimination, discussed in the next paragraphs.

STIMULUS DISCRIMINABILITY AND TRANSFER

Some studies have indicated that transfer may result from improvement in discriminability of stimuli following training. The hypotheses of these studies involve two different kinds of assumed processes. The first kind, represented in the papers of Gibson (17, 19) and Gibson and Gibson (20), assumes that differential reinforcement during learning serves to reduce the gradient of stimulus generalization. The second kind, which appears in the formulation of Miller and Dollard (33), assumes that the stimuli become more "distinctive" as a result of the addition of response-produced cues generated during learning. Both of these hypotheses involve the assumption that the operations of a learning experiment provide for development of increased stimulus discriminability. Both also postulate that discriminability thus developed persists, and is effective in a transfer task which involves differential response. Goss (21) has extended these postulates to include response relations in the two tasks. He assumes that responses made during the first task give rise to stimuli which, added to the task stimuli, make them more distinctive. He also assumes that these response-produced stimuli may be conditioned to the response of the transfer task, in addition to or independent of other conditioning. According to this view, if an instrumental response has been conditioned to a stimulus in the past, and if that stimulus is aroused by a response made in a learning task, transfer to a new task will occur by way of the relation between the response-produced stimulus and the instrumental response of the transfer task. Thus, in the experiment by Gagné and Baker, cited earlier (15), the transfer found could be accounted for on the basis that the labels learned in the first task facilitated establishment of distinctiveness by becoming part of the stimulus-complex employed in the second task. The positive transfer found by McAllister (30) could result from the pre-established relation between the label learned in the first task and the motor response of the second.

These principles may also be employed to account in part for the data of recognition studies. Added distinctiveness would reduce generalization and thus reduce the possibility of false recognitions in the second task. Pre-existing relations between responses made to the stimuli would facilitate appropriate instrumental responses which reflect recognition. In the experiment by Arnoult, for example, the addition of the label during paired-associates training could be regarded as facilitating discrimination of the old stimuli from the new, and this improved discrimination would reduce the likelihood of selection of a new form. In McAllister's study, increased relatedness of the response label to the response of switch selection, and the corresponding increase in transfer, could be regarded as reflecting correct recognition.

Here, as in the previous discussion, some qualification is needed. In terms of the paradigms cited earlier, recognition is conceived as depending upon differential response to discriminative stimuli. Pressing a correct switch when an old stimulus is presented is facilitated, presumably, by recognition resulting in part from the response of labeling. Pressing an incorrect switch may reflect failure of recognition and result in negative transfer. Conversely, as in the previous examples, recognition could result from positive transfer of a labeling response learned during the first task, and failure of recognition could result similarly. The interpretation of a recognition experiment requires careful analysis of the S–R relations involved if they are to be used to account for the results.

VERBAL MEDIATION AND TRANSFER

Several studies have suggested that transfer may be mediated by verbal associations, established either between the stimuli of the first task and the responses of the second or between the responses of the first task and the responses of the second.

Noble (34) and Noble and McNeely (35) have shown that verbal stimuli (words, syllables, or paralogs) are learned more rapidly if they also have greater numbers of associations made to them by independent groups of subjects. Postman (37) has also emphasized the view that associations made to stimuli lend meaning to them and facilitate transfer and improve discriminability. This latter view is taken by Postman as an alternative hypothesis, concerning the development of discrimination processes, to that of Gibson and Gibson (20). As such, the associationistic hypothesis represents an additional one which may apply to recognition as well. Facilitation of transfer would result, in this view, from enrichment of the stimuli, by way of the responses made to them during the first task. Recognition would also, presumably, be improved by similar enrichment. Stated another way, the facilitation could result from an increase in the probability of occurrence of a mediating association between the stimuli of the first task and the responses of the second. In Arnoult's experiment, for example, the improved recognition of the forms for which the subjects invented their own labels may have come about as a result of the use of labels having increased meaningfulness (2) as well as belongingness, as postulated (2).

In general, according to this view, responses to old stimuli would be facilitated by virtue of associations formed during learning, and responses to new stimuli would occur only as a result of generalization from old stimuli, on the basis of similarity, as in Osgood's model (36). Similarly, if associations were formed between the responses of the learning and transfer tasks, transfer and recognition would be facilitated in a manner such as that postulated by Goss (21).

IMPLICATIONS FOR THE STUDY OF RECOGNITION

The formulations of transfer that we have outlined above may be seen to involve four kinds of assumptions which are generalizable to the study of recognition:

1. Transfer depends on the degree to which responses learned to old stimuli generalize to new stimuli. This assumption is evident in the empirical model of Osgood (36). Recognition, similarly conceived, would depend upon the discriminability of the stimuli, but would be inversely related to transfer. Selection responses, made to new stimuli, would reflect positive transfer, but also failure of discrimination and false recognition. As a working hypothesis, this proposition appears fairly obvious, but it is not easy to test in all cases. The question of specification of relevant stimulus dimensions against which to relate discriminability is particularly difficult to answer, and only a bare beginning has been made in the development of physical variables which describe complex stimuli. In the case of forms and patterns, for example, physical variables have not been clearly described in many studies, and invariant

properties of "form" which distinguish them from other "forms" have not been adequately investigated (4, 5, 28) independently of learning (5) or judgments (23). Alteration of structure (24), judgments of meaning (23), and similar variables have been used as measures of "similarity," but without a defined physical scale. The efforts of Attneave and Arnoult (5) to develop measures of forms are important advances in this direction, as are other similar attempts (28). But greater efforts are clearly required before the relation between discriminability (25) and recognition (and transfer, for that matter) may be clearly established.

2. Transfer depends on the extent to which the stimuli of the training task become differentiated or distinctive, either as a result of reduction of generalization (17) or as a result of addition of response-produced cues (21, 33). To the extent that recognition is dependent upon discriminability, by these hypotheses it too would be facilitated and related to transfer, as in Paragraph 1 above. The added hypothesis here, of course, is that discriminability may be improved by differential response training, and that it persists and facilitates recognition.

3. As a corollary to the first hypothesis above, we would add that transfer and recognition are dependent upon the degree to which response-produced stimuli are themselves discriminable or distinctive. Clearly, in Goss's formulation, if the response-produced stimuli are not discriminable, they will not add distinctiveness to the task stimuli. Here, as in the first hypothesis, the specification of physical measures of similarity of such stimuli needs to be made before adequate tests of this hypothesis may be conducted. More often than not, response-produced cues have been inferred, and more direct establishment of their existence and their measurement would constitute an important advance.

4. Transfer and recognition are dependent upon associative connections, formed during learning, and their relations both to the first-task stimuli and to the second-task responses. The probability of arousal of such associations and their relations to the second-task responses would need to be manipulated in systematic fashion in order to test this hypothesis. Although evidence exists to support the hypothesis in part, variation of both aspects (probability of occurrence and relation to the second-task response) would need to be carried out in the same experiment to obtain a clear view of the relative contribution of each. Birge, for example, found that a naming response, learned for nonsense pictures, facilitated transfer of a reaching response from one picture to another for which the same name was learned (9). Buzzotta (11) was able to alter the probability that an association would occur to a response word used in a first-task practice period, and this probability was related to transfer in a second task involving different responses. Vanderplas and Garvin[4] found that recognition was related to the degree to which the stimuli of a paired-associates task (nonsense forms) elicited associations in an independent group used to scale the forms for association value. Independent variation of these variables in the same group of subjects would extend knowledge about their importance to a considerable degree.

The above hypotheses are not regarded as mutually exclusive, in the sense that if one is true others are not. Rather, they may be conceived as representing stages in the development of transfer or recognition as a result of first-task practice. To the extent that one begins with discriminable stimuli, differential responses may be made to different stimuli, thus (perhaps) increasing their discriminability. Associations made to the stimuli or responses may add to the likelihood of transfer, discrimination, and recognition. Differential response-produced cues, themselves discriminable, may further add to these processes, by way of their likelihood of occurrence or their likelihood of evoking instrumental responses upon which recognition depends. The relations among the hypotheses are further questions to be answered empirically.

REFERENCES

1. ARNOULT, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401–409.
2. ARNOULT, M. D. Recognition of shapes following paired-associates pretraining. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research.* Washington: Nat. Acad. Sci.-Nat. Res. Council Publ. No. 455, 1956, 1–9.
3. ARNOULT, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.* 1957, **54**, 339–350.
4. ATTNEAVE, F. Physical determinants of the judged complexity of shapes. *J. exp. Psychol.*, 1957, **53**, 221–227.
5. ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.
6. BAKER, K. E., & WYLIE, R. C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 623–638.
7. BALDWIN, R. D. Discrimination learning as a function of stimulus predifferentiation and mediated association training. Unpublished doctoral dissertation. State Univer. of Iowa, 1954.
8. BAUM, M. H. A study of concept attainment and verbal learning. Unpublished doctoral dissertation. Yale Univer., 1951.
9. BIRGE, J. S. The role of verbal response in transfer. Unpublished doctoral dissertation. Yale Univer., 1941.
10. BUGELSKI, B. R., & CADWALLADER, T. C. A reappraisal of the transfer and retroaction surface. *J. exp. Psychol.*, 1956, **52**, 360–366.
11. BUZZOTTA, V. R. Association-probability in the study of verbal behavior. Unpublished doctoral dissertation. Washington Univer., 1956.
12. CAMPBELL, V., & FREEMAN, J. T. Some functions of experimentally-induced language in perceptual learning. *Percept. mot. Skills*, 1955, **1**, 71–79.
13. CANTOR, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49**, 339–342.
14. COFER, C. N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
15. GAGNÉ, R. M., & BAKER, K. E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
16. GAGNÉ, R. M., BAKER, K. E., & FOSTER, H. Transfer of discrimination training to a motor task. *J. exp. Psychol.*, 1950, **40**, 314–328.
17. GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
18. GIBSON, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
19. GIBSON, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.
20. GIBSON, J. J., & GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–41.
21. GOSS, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.
22. GRAHAM, C. H. Sensation and perception in an objective psychology. *Psychol. Rev.*, 1958, **65**, 65–76.
23. HAAGEN, C. H. Synonymity, vividness, familiarity and association ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.
24. HAMILTON, R. J. Retroactive facilitation as a function of degree of similarity between tasks. *J. exp. Psychol.*, 1943, **32**, 363–376.
25. HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

26. KANFER, F. H. Perception: Identification and instrumental activity. *Psychol. Rev.*, 1956, **63**, 317–329.

27. KURTZ, K. H., & HOVLAND, C. I. The effect of verbalizations during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157–164.

28. LABERGE, D. L., & LAWRENCE, D. H. A method of generating visual forms of graded similarity. *Amer. Psychologist*, 1955, **10**, 401. (Abstract)

29. LAWRENCE, D. H. Transfer between discriminations involving the same cues but different responses. *Amer. Psychologist*, 1947, **2**, 303–304.

30. MCALLISTER, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.

31. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* (2nd. ed.) New York: Longmans, Green, 1952.

32. MCKINNEY, F., & MCGEOCH, J. A. The character and extent of transfer in retroactive inhibition: Disparate serial lists. *Amer. J. Psychol.*, 1935, **47**, 409–423.

33. MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

34. NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.

35. NOBLE, C. E. & MCNEELY, D. A. The role of meaningfulness (m) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

36. OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

37. POSTMAN, L. Association theory and perceptual learning. *Psychol. Rev.*, 1955, **62**, 438–446.

38. ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–115.

39. ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: The role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.

40. WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1955.

(Received March 21, 1958)

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN[1]

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

1. How is information about the physical world sensed, or detected, by the biological system?
2. In what form is information stored, or remembered?
3. How does information contained in storage, or in memory, influence recognition and behavior?

The first of these questions is in the province of sensory physiology, and is the only one for which appreciable understanding has been achieved. This article will be concerned primarily with the second and third questions, which are still subject to a vast amount of speculation, and where the few relevant facts currently supplied by neurophysiology have not yet been integrated into an acceptable theory.

With regard to the second question, two alternative positions have been maintained. The first suggests that storage of sensory information is in the form of coded representations or images, with some sort of one-to-one mapping between the sensory stimulus and the stored pattern. According to this hypothesis, if one understood the code or "wiring diagram" of the nervous system, one should, in principle, be able to discover exactly what an organism remembers by reconstructing the original sensory patterns from the "memory traces" which they have left, much as we might develop a photographic negative, or translate the pattern of electrical charges in the "memory" of a digital computer. This hypothesis is appealing in its simplicity and ready intelligibility, and a large family of theoretical brain models has been developed around the idea of a coded, representational memory (**2, 3, 9, 14**). The alternative approach, which stems from the tradition of British empiricism, hazards the guess that the images of stimuli may never really be recorded at all, and that the central nervous system simply acts as an intricate switching network, where retention takes the form of new connections, or pathways, between centers of activity. In many of the more recent developments of this position (Hebb's "cell assembly," and Hull's "cortical anticipatory goal response," for example) the "responses" which are associated to stimuli may be entirely contained within the CNS itself. In this case the response represents an "idea" rather than an action. The important feature of this approach is that there is never any simple mapping of the stimulus into memory, according to some code which would permit its later reconstruction. Whatever information is retained must somehow be stored as a *preference for a particular response;* i.e., the information is contained in *connections* or *associations* rather than topographic representations. (The term *response*, for the remainder of this presentation, should be understood to mean any distinguishable state of the organism, which may or may not involve externally detectable muscular activity. The activation of some nucleus of cells in the central nervous system, for example, can constitute a response, according to this definition.)

---

[1] The development of this theory has been carried out at the Cornell Aeronautical Laboratory, Inc., under the sponsorship of the Office of Naval Research, Contract Nonr-2381(00). This article is primarily an adaptation of material reported in Ref. 15, which constitutes the first full report on the program.

Corresponding to these two positions on the method of information retention, there exist two hypotheses with regard to the third question, the manner in which stored information exerts its influence on current activity. The "coded memory theorists" are forced to conclude that recognition of any stimulus involves the matching or systematic comparison of the contents of storage with incoming sensory patterns, in order to determine whether the current stimulus has been seen before, and to determine the appropriate response from the organism. The theorists in the empiricist tradition, on the other hand, have essentially combined the answer to the third question with their answer to the second: since the stored information takes the form of new connections, or transmission channels in the nervous system (or the creation of conditions which are functionally equivalent to new connections), it follows that the new stimuli will make use of these new pathways which have been created, automatically activating the appropriate response without requiring any separate process for their recognition or identification.

The theory to be presented here takes the empiricist, or "connectionist" position with regard to these questions. The theory has been developed for a hypothetical nervous system, or machine, called a *perceptron*. The perceptron is designed to illustrate some of the fundamental properties of intelligent systems in general, without becoming too deeply enmeshed in the special, and frequently unknown, conditions which hold for particular biological organisms. The analogy between the perceptron and biological systems should be readily apparent to the reader.

During the last few decades, the development of symbolic logic, digital computers, and switching theory has impressed many theorists with the functional similarity between a neuron and the simple on-off units of which computers are constructed, and has provided the analytical methods necessary for representing highly complex logical functions in terms of such elements. The result has been a profusion of brain models which amount simply to logical contrivances for performing particular algorithms (representing "recall," stimulus comparison, transformation, and various kinds of analysis) in response to sequences of stimuli—e.g., Rashevsky (**14**), McCulloch (**10**), McCulloch & Pitts (**11**), Culbertson (**2**), Kleene (**8**), and Minsky (**13**). A relatively small number of theorists, like Ashby (**1**) and von Neumann (**17**, **18**), have been concerned with the problems of how an imperfect neural network, containing many random connections, can be made to perform reliably those functions which might be represented by idealized wiring diagrams. Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations. The need for a suitable language for the mathematical analysis of events in systems where only the gross organization can be characterized, and the precise structure is unknown, has led the author to formulate the current model in terms of probability theory rather than symbolic logic.

The theorists referred to above were chiefly concerned with the question of how such functions as perception and recall might be achieved by a deterministic physical system of any sort, rather than how this is actually done by the brain. The models which have been produced all fail in some important respects (absence of equipotentiality, lack of neuroeconomy, excessive specificity of connections and synchronization requirements, unrealistic specificity of stimuli sufficient for cell firing, postulation of variables or functional features with no known neurological correlates, etc.) to correspond to a biological system. The proponents of this line of approach have maintained that, once it has been shown how a physical system of any variety might be made to perceive and recognize stimuli, or perform other brainlike functions, it would require only a refinement or modification of existing principles to understand the working of a more realistic nervous system, and to eliminate the shortcomings mentioned above. The writer takes the position, on the other hand, that these shortcomings are such that a mere refinement or improvement of the principles already suggested can never account for biological intelligence; a *difference in principle* is clearly indicated. The theory of statistical separability (Cf. **15**), which is to be summarized here, appears to offer a solution in principle to all of these difficulties.

Those theorists—Hebb (**7**), Milner (**12**), Eccles (**4**), Hayek (**6**)—who have been more directly concerned with the biological nervous system and its activity in a natural environment, rather than with formally analogous machines, have generally been less exact in their formulations and far from rigorous in their analysis, so that it is frequently hard to assess whether or not the systems that they describe could actually work in a realistic nervous system, and what the necessary and sufficient conditions might be. Here again, the lack of an analytic language comparable in proficiency to the Boolean algebra of the network analysts has been one of the main obstacles. The contributions of this group should perhaps be considered as suggestions of what to look for and investigate, rather than as finished theoretical systems in their own right. Seen from this viewpoint, the most suggestive work, from the standpoint of the following theory, is that of Hebb and Hayek.

The position, elaborated by Hebb (**7**), Hayek (**6**), Uttley (**16**), and Ashby (**1**), in particular, upon which the theory of the perceptron is based, can be summarized by the following assumptions:

1. The physical connections of the nervous system which are involved in learning and recognition are not identical from one organism to another. At birth, the construction of the most important networks is largely random, subject to a minimum number of genetic constraints.
2. The original system of connected cells is capable of a certain amount of plasticity; after a period of neural activity, the probability that a stimulus applied to one set of cells will cause a response in some other set is likely to change, due to some relatively long-lasting changes in the neurons themselves.
3. Through exposure to a large sample of stimuli, those which are most "similar" (in some sense which must be defined in terms of the particular physical system) will tend

to form pathways to the same sets of responding cells. Those which are markedly "dissimilar" will tend to develop connections to different sets of responding cells.

4. The application of positive and/or negative reinforcement (or stimuli which serve this function) may facilitate or hinder whatever formation of connections is currently in progress.

5. *Similarity*, in such a system, is represented at some level of the nervous system by a tendency of similar stimuli to activate the same sets of cells. Similarity is not a necessary attribute of particular formal or geometrical classes of stimuli, but depends on the physical organization of the perceiving system, an organization which evolves through interaction with a given environment. The structure of the system, as well as the ecology of the stimulus-environment, will affect, and will largely determine, the classes of "things" into which the perceptual world is divided.

THE ORGANIZATION OF A PERCEPTRON

The organization of a typical photoperceptron (a perceptron responding to optical patterns as stimuli) is shown in Fig. 1. The rules of its organization are as follows:

1. Stimuli impinge on a retina of sensory units (S-points), which are assumed to respond on an all-or-nothing basis, in some models, or with a pulse amplitude or frequency proportional to the stimulus intensity, in other models. In the models considered here, an all-or-nothing response will be assumed.

2. Impulses are transmitted to a set of association cells (A-units) in a "projection area" (A_I). This projection area may be omitted in some models, where the retina is connected directly to the association area (A_{II}). The cells in the projection area each receive a number of connections from the sensory points. The set of S-points transmitting impulses to a particular A-unit will be called the *origin points* of that A-unit. These origin points may be either *excitatory* or *inhibitory* in their effect on the A-unit. If the algebraic sum of excitatory and inhibitory impulse intensities is equal to or greater than the threshold (θ) of the A-unit, then the A-unit fires, again on an all-or-nothing basis (or, in some models, which will not be considered here, with a frequency which depends on the net value of the impulses received). The origin points of the A-units in the projection area tend to be clustered or focalized, about some central point, corresponding to each A-unit. The number of origin points falls off exponentially as the retinal distance from the central point for the A-unit in question increases. (Such a distribution seems to be supported by physiological evidence, and serves an important functional purpose in contour detection.)

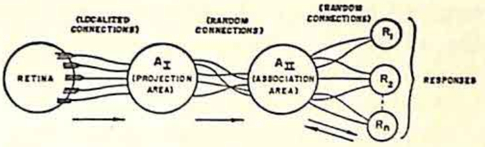


FIG. 1. Organization of a perceptron.

3. Between the projection area and the association area (A_{II}), connections are assumed to be random. That is, each A-unit in the A_{II} set receives some number of fibers from origin points in the A_I set, but these origin points are scattered at random throughout the projection area. Apart from their connection distribution, the A_{II} units are identical with the A_I units, and respond under similar conditions.

4. The "responses," $R_1, R_2, \ldots, R_n$ are cells (or sets of cells) which

respond in much the same fashion as the A-units. Each response has a typically large number of origin points located at random in the A_{II} set. The set of A-units transmitting impulses to a particular response will be called the source-set for that response. (The source-set of a response is identical to its set of origin points in the A-system.) The arrows in Fig. 1 indicate the direction of transmission through the network. Note that up to A_{II} all connections are forward, and there is no feedback. When we come to the last set of connections, between A_{II} and the R-units, connections are established in both directions. The rule governing feedback connections, in most models of the perceptron, can be either of the following alternatives:

(*a*) Each response has excitatory feedback connections to the cells in its own source-set, or

(*b*) Each response has inhibitory feedback connections to the complement of its own source-set (i.e., it tends to prohibit activity in any association cells which do not transmit to it).

The first of these rules seems more plausible anatomically, since the R-units might be located in the same cortical area as their respective source-sets, making mutual excitation between the R-units and the A-units of the appropriate source-set highly probable. The alternative rule (*b*) leads to a more readily analyzed system, however, and will therefore be assumed for most of the systems to be evaluated here.

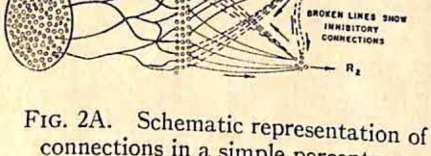


FIG. 2A. Schematic representation of connections in a simple perceptron.

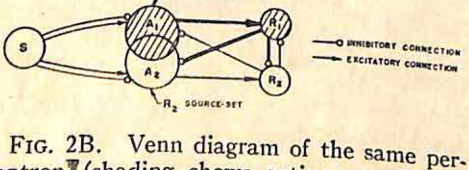


FIG. 2B. Venn diagram of the same perceptron (shading shows active sets for R_1 response).

Figure 2 shows the organization of a simplified perceptron, which affords a convenient entry into the theory of statistical separability. After the theory has been developed for this simplified model, we will be in a better position to discuss the advantages of the system in Fig. 1. The feedback connections shown in Fig. 2 are inhibitory, and go to the complement of the source-set for the response from which they originate; consequently, this system is organized according to Rule *b*, above. The system shown here has only three stages, the first association stage having been eliminated. Each A-unit has a set of randomly located origin points in the retina. Such a system will form similarity concepts on the basis of *coincident areas* of stimuli, rather than by the similarity of contours or outlines. While such a system is at a disadvantage in many discrimination experiments, its capability is still quite impressive, as will be demonstrated presently. The system shown in Fig. 2 has only two responses, but there is clearly no limit on the number that might be included.

The responses in a system organized in this fashion are mutually exclusive. If R_1 occurs, it will tend to inhibit R_2, and will also inhibit the source-set for R_2. Likewise, if R_2 should occur, it will tend to inhibit R_1. If the total impulse received from all the A-units in one source-set is stronger or more frequent than the impulse received by the alternative (antagonistic) response, then the first response will tend to gain an advantage over the other, and will be the one which occurs. If such a system is to be capable of learning, then it must be possible to modify the A-units or their connections in such a way that stimuli of one class will tend to evoke a stronger impulse in the R_1 source-set than in the R_2 source-set, while stimuli of another (dissimilar) class will tend to evoke a stronger impulse in the R_2 source-set than in the R_1 source-set.

It will be assumed that the impulses delivered by each A-unit can be characterized by a value, V, which may be an amplitude, frequency, latency, or probability of completing transmission. If an A-unit has a high value, then all of its output impulses are considered to be more effective, more potent, or more likely to arrive at their endbulbs than impulses from an A-unit with a lower value. The value of an A-unit is considered to be a fairly stable characteristic, probably depending on the metabolic condition of the cell and the cell membrane, but it is not absolutely constant. It is assumed that, in general, periods of activity tend to increase a cell's value, while the value may decay (in some models) with inactivity. The most interesting models are those in which cells are assumed to compete for metabolic materials, the more active cells gaining at the expense of the less active cells. In such a system, if there is no activity, all cells will tend to remain in a relatively constant condition, and (regardless of activity) the net value of the system, taken in

TABLE 1

COMPARISON OF LOGICAL CHARACTERISTICS OF α, β, AND γ SYSTEMS

	α-System (Uncompensated Gain System)	β-System (Constant Feed System)	γ-System (Parasitic Gain System)
Total value-gain of source set per reinforcement	N_{a_r}	K	0
ΔV for A-units active for 1 unit of time	+1	K/N_{a_r}	+1
ΔV for inactive A-units outside of dominant set	0	K/N_{A_r}	0
ΔV for inactive A-units of dominant set	0	0	$\frac{-N_{a_r}}{N_{A_r}-N_{a_r}}$
Mean value of A-system	Increases with number of reinforcements	Increases with time	Constant
Difference between mean values of source-sets	Proportional to differences of reinforcement frequency $(n_{s_{r1}}-n_{s_{r2}})$	0	0

Note: In the β and γ systems, the total value-change for any A-unit will be the sum of the ΔV's for all source-sets of which it is a member.

N_{a_r} = Number of active units in source-set
N_{A_r} = Total number of units in source-set
$n_{s_{rj}}$ = Number of stimuli associated to response r_j
K = Arbitrary constant

its entirety, will remain constant at all times. Three types of systems, which differ in their value dynamics, have been investigated quantitatively. Their principal logical features are compared in Table 1. In the alpha system, an active cell simply gains an increment of value for every impulse, and holds this gain indefinitely. In the beta system, each source-set is allowed a certain constant rate of gain, the increments being apportioned among the cells of the source-set in proportion to their activity. In the gamma system, active cells gain in value at the expense of the inactive cells of their source-set, so that the total value of a source-set is always constant.

For purposes of analysis, it is convenient to distinguish two phases in the response of the system to a stimulus (Fig. 3). In the *predominant phase*, some proportion of A-units (represented by solid dots in the figure) responds to the stimulus, but the R-units are still inactive. This phase is transient, and quickly gives way to the *postdominant phase*, in which one of the responses becomes active, inhibiting activity in the complement of its own source-set, and thus preventing the occurrence of any alternative response. The response which happens to become dominant is initially random, but if the A-units are reinforced (i.e., if the active units are allowed to gain in value), then when the same stimulus is presented again at a later time, the same response will have a stronger tendency to recur, and learning can be said to have taken place.

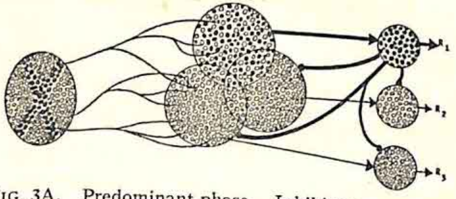


FIG. 3A. Predominant phase. Inhibitory connections are not shown. Solid black units are active.

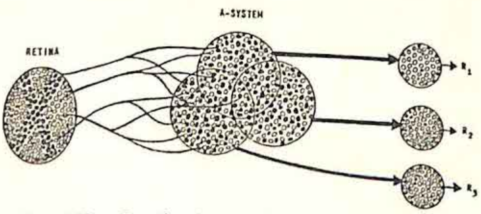


FIG. 3B. Postdominant phase. Dominant subset suppresses rival sets. Inhibitory connections shown only for R1.

FIG. 3. Phases of response to a stimulus.

ANALYSIS OF THE PREDOMINANT PHASE

The perceptrons considered here will always assume a fixed threshold, θ, for the activation of the A-units. Such a system will be called a *fixed-threshold model*, in contrast to a *continuous transducer model*, where the response of the A-unit is some continuous function of the impinging stimulus energy.

In order to predict the learning curves of a fixed-threshold perceptron, two variables have been found to be of primary importance. They are defined as follows:

P_a = the expected proportion of A-units activated by a stimulus of a given size,

P_c = the conditional probability that an A-unit which responds to a given stimulus, S_1, will also respond to another given stimulus, S_2.

It can be shown (Rosenblatt, **15**) that as the size of the retina is increased, the number of S-points (N_s) quickly ceases to be a significant parameter, and the values of P_a and P_c approach the value that they would have for a retina with infinitely many points. For a large retina, therefore, the equations are as follows:

$$P_a = \sum_{e=\theta}^{x} \sum_{i=\theta}^{\substack{min \\ (y,\, e-\theta)}} P(e,i) \qquad (1)$$

where

$$P(e,i) = \binom{x}{e} R^{e}(1-R)^{x-e} \times \binom{y}{i} R^{i}(1-R)^{y-i}$$

and

R = proportion of S-points activated by the stimulus

x = number of excitatory connections to each A-unit

y = number of inhibitory connections to each A-unit

θ = threshold of A-units.

(The quantities e and i are the excitatory and inhibitory components of the excitation received by the A-unit from the stimulus. If the algebraic sum $\alpha = e + i$ is equal to or greater than θ, the A-unit is assumed to respond.)

$$P_c = \frac{1}{P_a} \sum_{e=\theta}^{x} \sum_{i=e-\theta}^{y} \sum_{l_e=0}^{e} \sum_{l_i=0}^{i} \sum_{g_e=0}^{x-e} \sum_{g_i=0}^{y-i} P(e,i,l_e,l_i,g_e,g_i) \qquad (2)$$

$$(e - i - l_e + l_i + g_e - g_i \geq \theta)$$

where

$$\begin{aligned} P(e,i,l_e,l_i,g_e,g_i) &= \binom{x}{e} R^{e}(1-R)^{x-e} \\ &\times \binom{y}{i} R^{i}(1-R)^{y-i} \\ &\times \binom{e}{l_e} L^{l_e}(1-L)^{e-l_e} \\ &\times \binom{i}{l_i} L^{l_i}(1-L)^{i-l_i} \\ &\times \binom{x-e}{g_e} G^{g_e}(1-G)^{x-e-g_e} \\ &\times \binom{y-i}{g_i} G^{g_i}(1-G)^{y-i-g_i} \end{aligned}$$

and

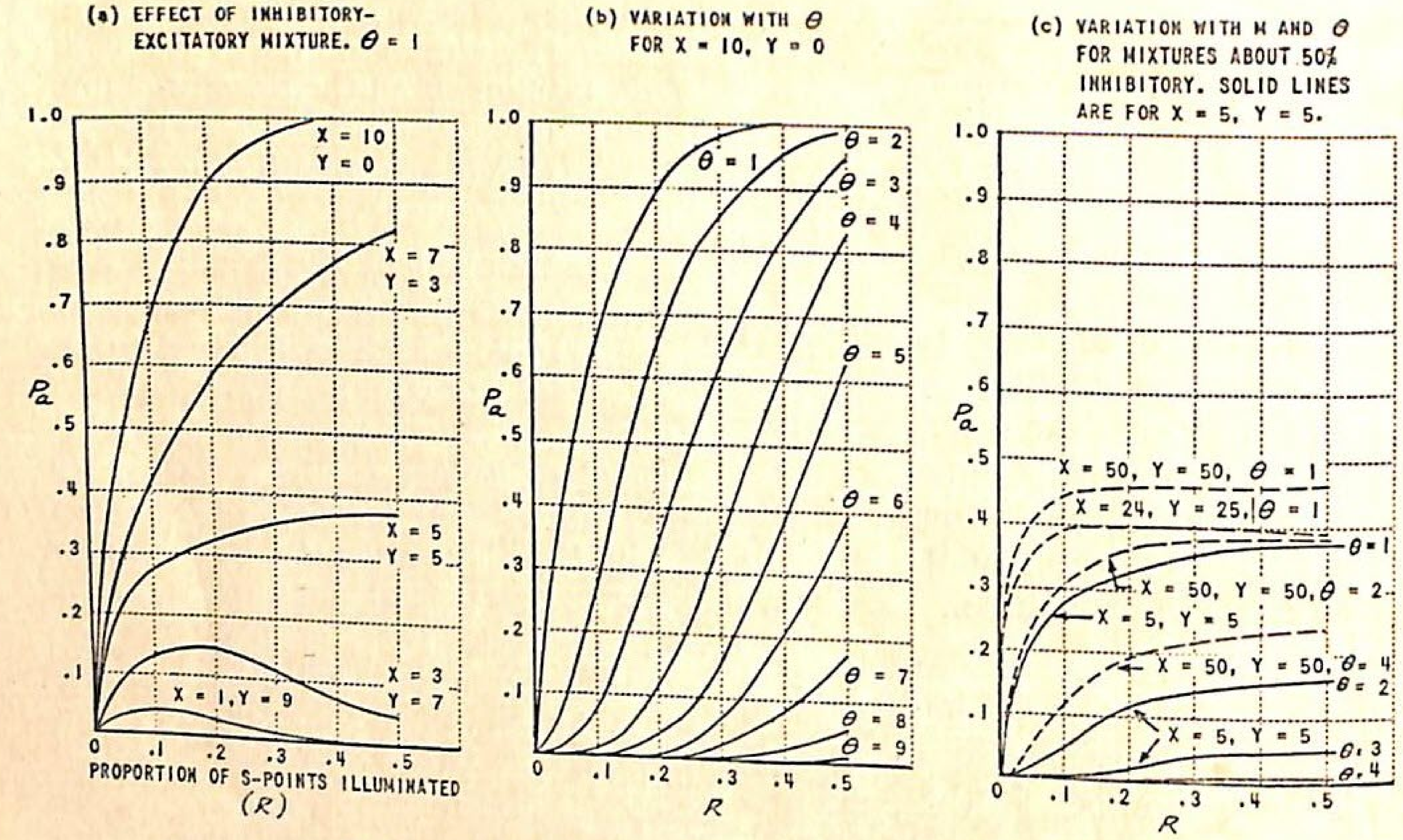
L = proportion of the S-points illuminated by the first stimulus, S_1, which are not illuminated by S_2

G = proportion of the residual S-set (left over from the first stimulus) which is included in the second stimulus (S_2).

The quantities R, L, and G specify the two stimuli and their retinal overlap. l_e and l_i are, respectively, the numbers of excitatory and inhibitory origin points "lost" by the A-unit when stimulus S_1 is replaced by S_2; g_e and g_i are the numbers of excitatory and inhibitory origin points "gained" when stimulus S_1 is replaced by S_2. The summations in Equation 2 are between the limits indicated, subject to the side condition $e - i - l_e + l_i + g_e - g_i \geq \theta$.

Some of the most important characteristics of P_a are illustrated in Fig. 4, which shows P_a as a function of the retinal area illuminated (R). Note that P_a can be reduced in magnitude by either increasing the threshold, θ, or by increasing the proportion of inhibitory connections (y). A comparison of Fig. 4b and 4c shows that if the excitation is about equal to the inhibition, the curves for P_a as a function of R are flattened out, so that there is little variation in P_a for stimuli of different sizes. This fact is of great importance for systems which require P_a to be close to an optimum value in order to perform properly.

The behavior of P_c is illustrated in Fig. 5 and 6. The curves in Fig. 5 can be compared with those for P_a in Fig. 4. Note that as the threshold is increased, there is an even sharper reduction in the value of P_c than was the case with P_a. P_c also decreases as the proportion of inhibitory connections increases, as does P_a. Fig. 5, which is



(a) EFFECT OF INHIBITORY-EXCITATORY MIXTURE. $\theta = 1$

(b) VARIATION WITH θ FOR X = 10, Y = 0

(c) VARIATION WITH N AND θ FOR MIXTURES ABOUT 50% INHIBITORY. SOLID LINES ARE FOR X = 5, Y = 5.

FIG. 4. P_a as function of retinal area illuminated.

calculated for nonoverlapping stimuli, illustrates the fact that P_c remains greater than zero even when the stimuli are completely disjunct, and illuminate no retinal points in common. In Fig. 6, the effect of varying amounts of overlap between the stimuli is shown. In all cases, the value of P_c goes to unity as the stimuli approach perfect identity. For smaller stimuli (broken line curves), the value of P_c is lower than for large stimuli. Similarly, the value is less for high thresholds than for low thresholds. The minimum value of P_c will be equal to

$$P_{c_{min}} = (1 - L)^x(1 - G)^y. \quad (3)$$

In Fig. 6, $P_{c_{min}}$ corresponds to the curve for $\theta = 10$. Note that under these conditions the probability that the A-unit responds to both stimuli (P_c) is practically zero, except for stimuli which are quite close to identity. This condition can be of considerable help in discrimination learning.

MATHEMATICAL ANALYSIS OF LEARNING IN THE PERCEPTRON

The response of the perceptron in the predominant phase, where some fraction of the A-units (scattered throughout the system) responds to the stimulus, quickly gives way to the postdominant response, in which activity is limited to a single source-set, the other sets being suppressed. Two possible systems have been studied for the determination of the "dominant" response, in the postdominant phase. In one (the mean-discriminating system, or μ-system), the response whose inputs have the greatest mean value responds first, gaining a slight advantage over the others, so that it quickly becomes dominant. In the second case (the sum-discriminating system, or Σ-system), the response whose inputs have the greatest net value gains an advantage. In most cases, systems which respond to mean values have an advantage over systems which respond to sums, since the means are

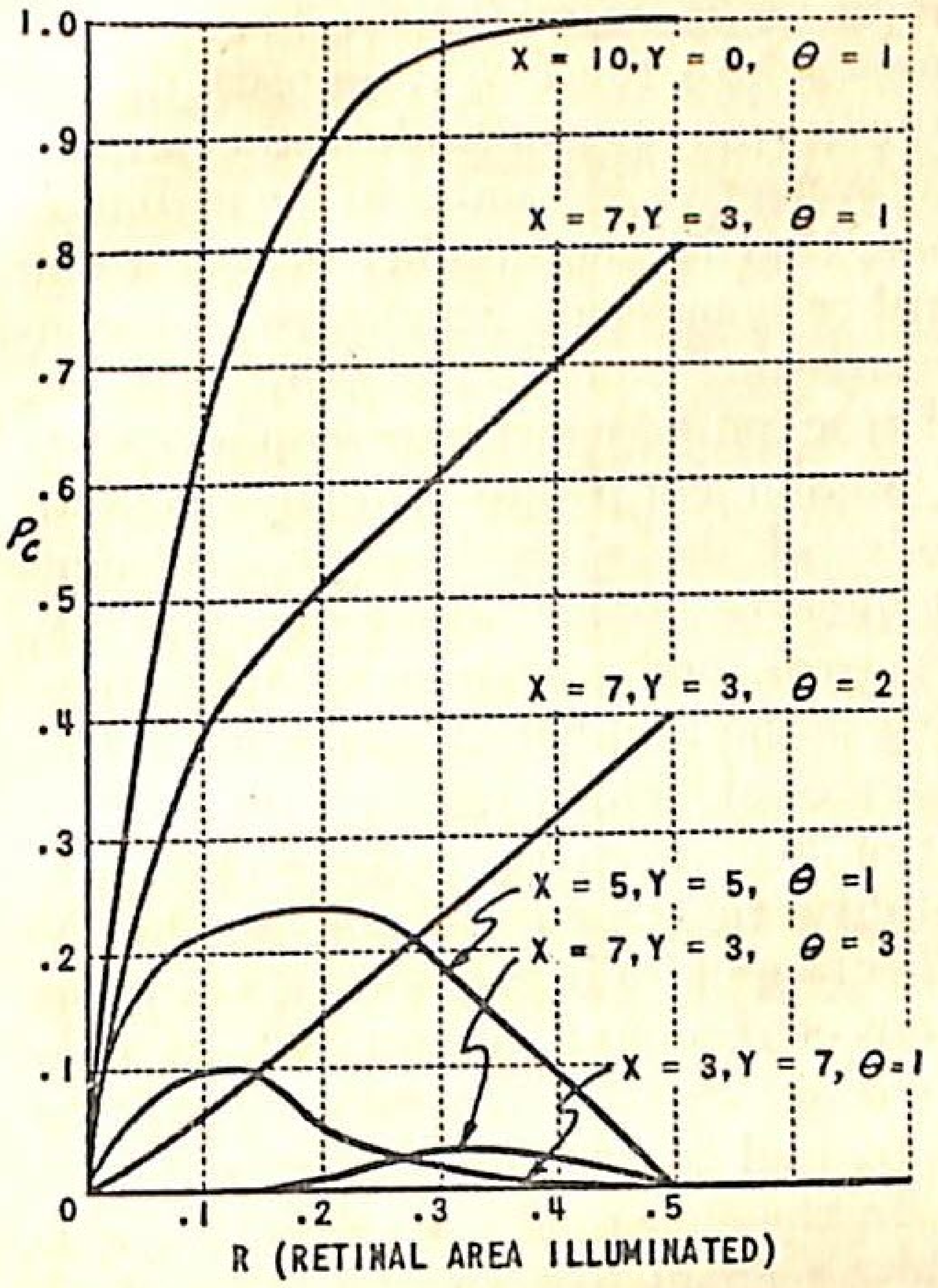


FIG. 5. P_c as a function of R, for nonoverlapping stimuli.

less influenced by random variations in P_a from one source-set to another. In the case of the γ-system (see Table 1), however, the performance of the μ-system and Σ-system become identical.

We have indicated that the perceptron is expected to learn, or to form associations, as a result of the changes in value that occur as a result of the activity of the association cells. In evaluating this learning, one of two types of hypothetical experiments can be considered. In the first case, the perceptron is exposed to some series of stimulus patterns (which might be presented in random positions on the retina) and is "forced" to give the desired response in each case. (This forcing of responses is assumed to be a prerogative of the experimenter. In experiments intended to evaluate trial-and-error learning, with more sophisticated perceptrons, the experimenter does not force the system to respond in the desired fashion, but merely applies positive reinforcement when the response happens to be correct, and negative reinforcement when the response is wrong.) In evaluating the learning which has taken place during this "learning series," the perceptron is assumed to be "frozen" in its current condition, no further value changes being allowed, and the same series of stimuli is presented again in precisely the same fashion, so that the stimuli fall on identical positions on the retina. The probability that the perceptron will show a bias towards the "correct" response (the one which has been previously reinforced during the learning series) in preference to any given alternative response is called P_r, the probability of correct choice of response between two alternatives.

In the second type of experiment, a learning series is presented exactly as before, but instead of evaluating the perceptron's performance using the same series of stimuli which were shown before, a new series is presented, in which stimuli may be drawn from the same *classes* that were previ-

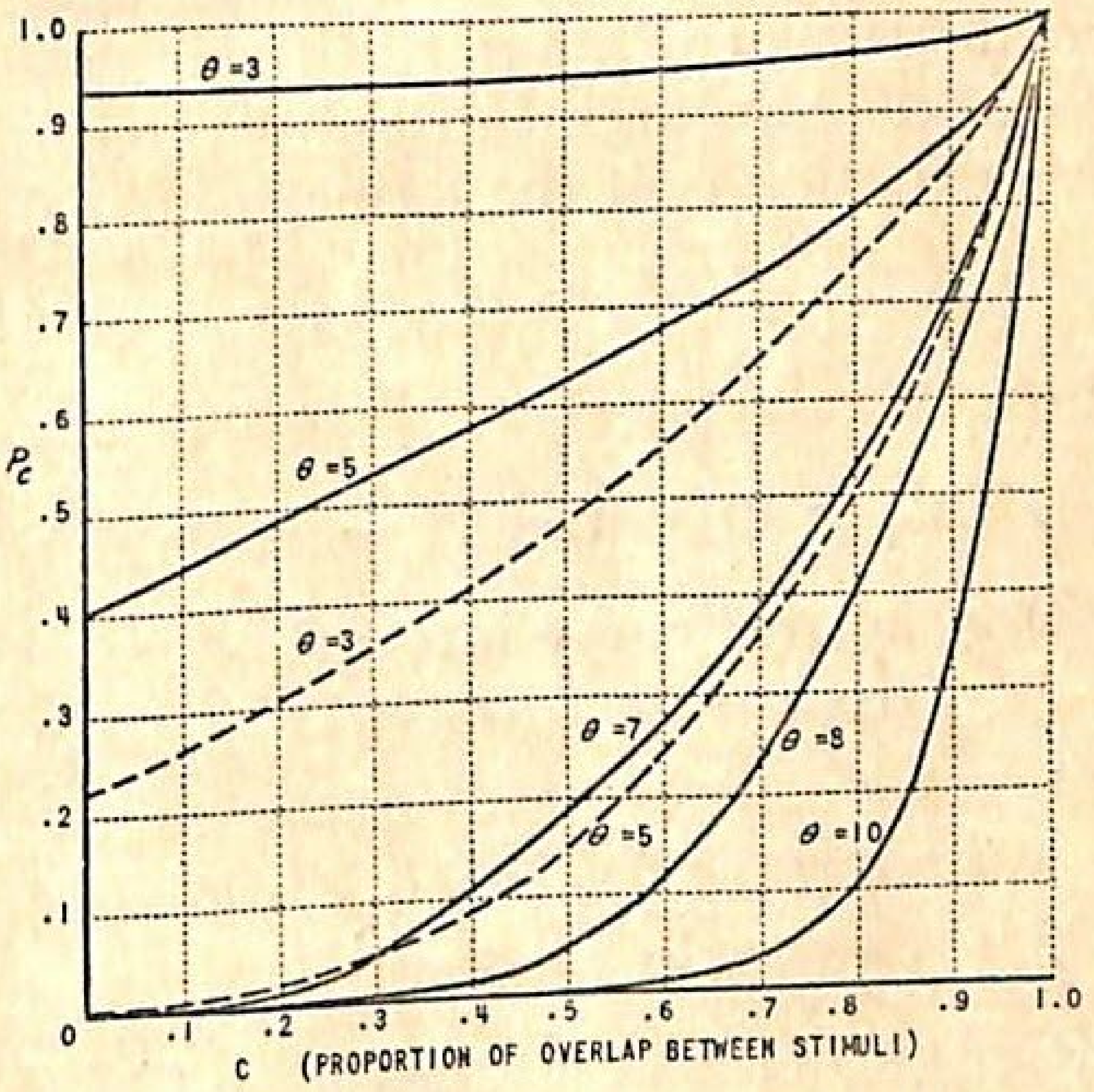


FIG. 6. P_c as a function of C. $X = 10$, $Y = 0$. Solid lines: R = .5; broken lines: R = .2.

ously experienced, but are not necessarily identical. This new test series is assumed to be composed of stimuli projected onto random retinal positions, which are chosen independently of the positions selected for the learning series. The stimuli of the test series may also differ in size or rotational position from the stimuli which were previously experienced. In this case, we are interested in the probability that the perceptron will give the correct response for the *class* of stimuli which is represented, regardless of whether the particular stimulus has been seen before or not. This probability is called P_g, the probability of correct generalization. As with P_r, P_g is actually the probability that a bias will be found in favor of the proper response rather than any one alternative; only one pair of responses at a time is considered, and the fact that the response bias is correct in one pair does not mean that there may not be other pairs in which the bias favors the wrong response. The probability that the correct response will be preferred over *all* alternatives is designated P_R or P_G.

In all cases investigated, a single general equation gives a close approximation to P_r and P_g, if the appropriate constants are substituted. This equation is of the form:

$$P = P(N_{a_r} > 0) \cdot \phi(Z) \qquad (4)$$

where

$$P(N_{a_r} > 0) = 1 - (1 - P_a)^{N_e}$$

$\phi(Z) =$ normal curve integral from $-\infty$ to Z

and

$$Z = \frac{c_1 n_{s_r} + c_2}{\sqrt{c_3 n_{s_r}{}^2 + c_4 n_{s_r}}}.$$

If R_1 is the "correct" response, and R_2 is the alternative response under consideration, Equation 4 is the probability that R_1 will be preferred over R_2 after n_{s_r} stimuli have been shown for each of the two responses, during the learning period. N_e is the number of "effective" A-units in each source-set; that is, the number of A-units in either source-set which are not connected in common to both responses. Those units which are connected in common contribute equally to both sides of the value balance, and consequently do not affect the net bias towards one response or the other. N_{a_r} is the number of active units in a source-set, which respond to the test stimulus, $S_t \cdot P(N_{a_r} > 0)$ is the probability that at least one of the N_e effective units in the source-set of the correct response (designated, by convention, as the R_1 response) will be activated by the test stimulus, S_t.

In the case of P_g, the constant c_2 is always equal to zero, the other three constants being the same as for P_r. The values of the four constants depend on the parameters of the physical nerve net (the perceptron) and also on the organization of the stimulus environment.

The simplest cases to analyze are those in which the perceptron is shown stimuli drawn from an "ideal environment," consisting of randomly placed points of illumination, where there is no attempt to classify stimuli according to intrinsic similarity. Thus, in a typical learning experiment, we might show the perceptron 1,000 stimuli made up of random collections of illuminated retinal points, and we might arbitrarily reinforce R_1 as the "correct" response for the first 500 of these, and R_2 for the remaining 500. This environment is "ideal" only in the sense that we speak of an ideal gas in physics; it is a convenient artifact for purposes of analysis, and does not lead to the best performance from the perceptron. In the ideal environment situation, the constant c_1 is always equal to zero, so that, in the case of P_g (where c_2 is also zero), the value of Z will be zero, and P_g can never be any better than the random expectation of 0.5. The evaluation of P_r for these conditions, however, throws some interesting light on the differences between the alpha, beta, and gamma systems (Table 1).

First consider the alpha system, which has the simplest dynamics of the three. In this system, whenever an A-unit is active for one unit of time, it gains one unit of value. We will assume an experiment, initially, in which N_{s_r} (the number of stimuli associated to each response) is constant for all responses. In this case, for the sum system,

$$\left.\begin{aligned} c_1 &= 0 \\ c_2 &= (1 - P_a)N_e \\ c_3 &= 2P_a\omega \\ c_4 &\approx 0 \end{aligned}\right\} \qquad (5)$$

where ω = the fraction of responses connected to each A-unit. If the source-sets are disjunct, $\omega = 1/N_R$, where N_R is the number of responses in the system. For the μ-system,

$$\left.\begin{aligned} c_1 &= 0 \\ c_2 &= (1 - P_a)N_e \\ c_3 &= 0 \\ c_4 &= 2\omega \end{aligned}\right\} \qquad (6)$$

The reduction of c_3 to zero gives the μ-system a definite advantage over the Σ-system. Typical learning curves for these systems are compared in Fig. 7 and 8. Figure 9 shows the effect of variations in P_a upon the performance of the system.

If n_{s_r}, instead of being fixed, is treated as a random variable, so that the number of stimuli associated to each response is drawn separately from some distribution, then the per-

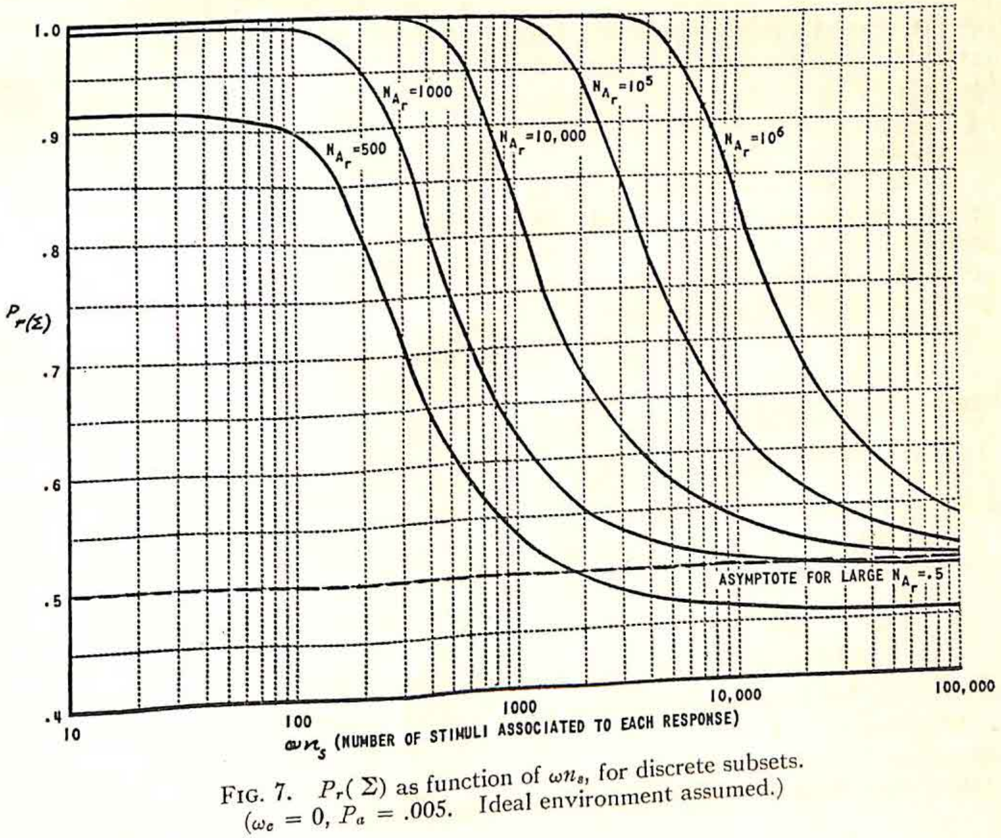


FIG. 7. $P_r(\Sigma)$ as function of ωn_s, for discrete subsets. ($\omega_c = 0$, $P_a = .005$. Ideal environment assumed.)

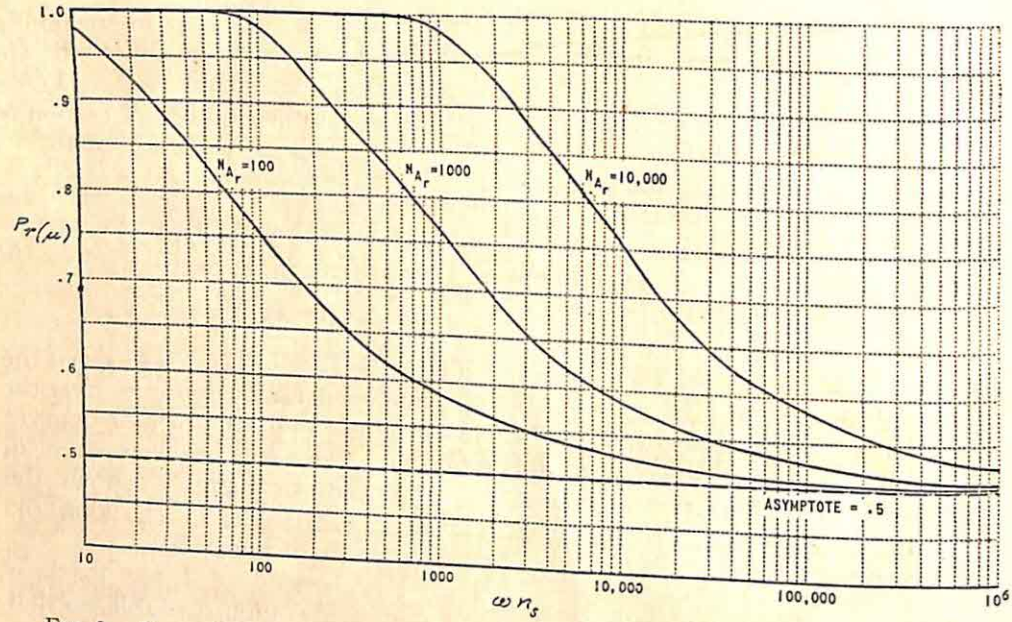


FIG. 8. $P_{r(\mu)}$ as function of ωn_s. (For $P_a = .07$, $\omega_c = 0$. Ideal environment assumed.)

formance of the α-system is considerably poorer than the above equations indicate. Under these conditions, the constants for the μ-system are

$$\left.\begin{aligned} c_1 &= 0 \\ c_2 &= 1 - P_a \\ c_3 &= 2P_a^2q^2\left[\frac{(\omega N_R - 1)^2}{N_R - 2} + 1\right] \\ c_4 &= \frac{2(1 - P_a)N_R}{(1 - \omega_c)N_A} \end{aligned}\right\} \quad (7)$$

where

q = ratio of $\sigma_{n_{s_r}}$ to $\bar{n}_{s_r}$
N_R = number of responses in the system
N_A = number of A-units in the system
ω_c = proportion of A-units common to R_1 and R_2.

For this equation (and any others in which n_{s_r} is treated as a random variable), it is necessary to define n_{s_r} in Equation 4 as the expected value of this variable, over the set of all responses.

For the β-system, there is an even greater deficit in performance, due to the fact that the net value continues to grow regardless of what happens to the system. The large net values of the subsets activated by a stimulus tend to amplify small statistical differences, causing an unreliable performance. The constants in this case (again for the μ-system) are

$$\left.\begin{aligned} c_1 &= 0 \\ c_2 &= (1 - P_a)N_e \\ c_3 &= 2(P_aN_eq\omega N_R^2)^2 \\ c_4 &= 2(1 - P_a)\omega N_RN_e \end{aligned}\right\} \quad (8)$$

In both the alpha and beta systems, performance will be poorer for the sum-discriminating model than for the mean-discriminating case. In the gamma-system, however, it can be shown that $P_{r(\Sigma)} = P_{r(\mu)}$; i.e., it makes no difference in performance whether the Σ-system or μ-system is used. Moreover, the constants for the γ-system, with variable n_{s_r}, are identical to the constants for the alpha μ-sys-

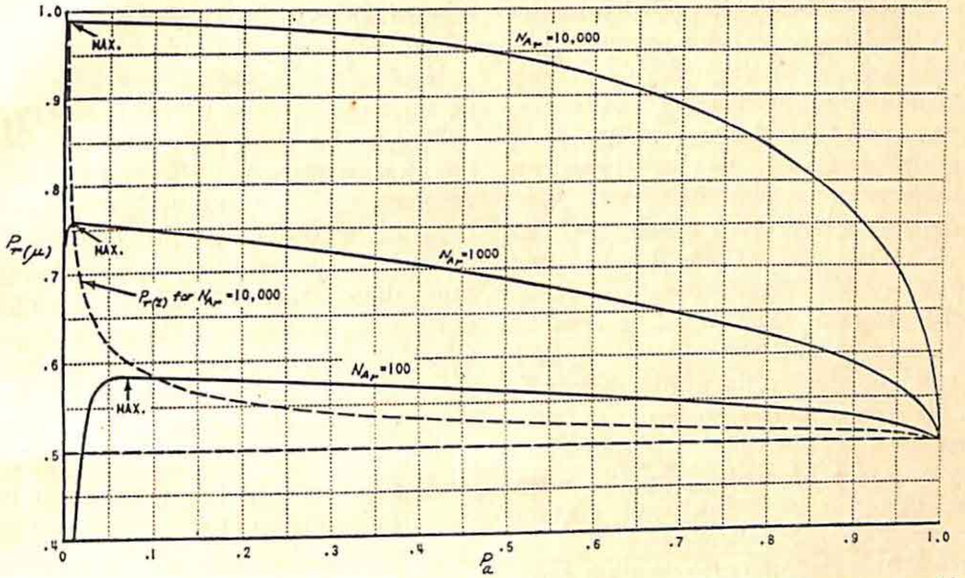


FIG. 9. $P_{r(\mu)}$ as function of P_a. (For $n_{s_r} = 1{,}000$, $\omega_c = 0$. Ideal environment assumed.)

tem, with n_{s_r} fixed (Equation 6). The performance of the three systems is compared in Fig. 10, which clearly demonstrates the advantage of the γ-system.

Let us now replace the "ideal en-

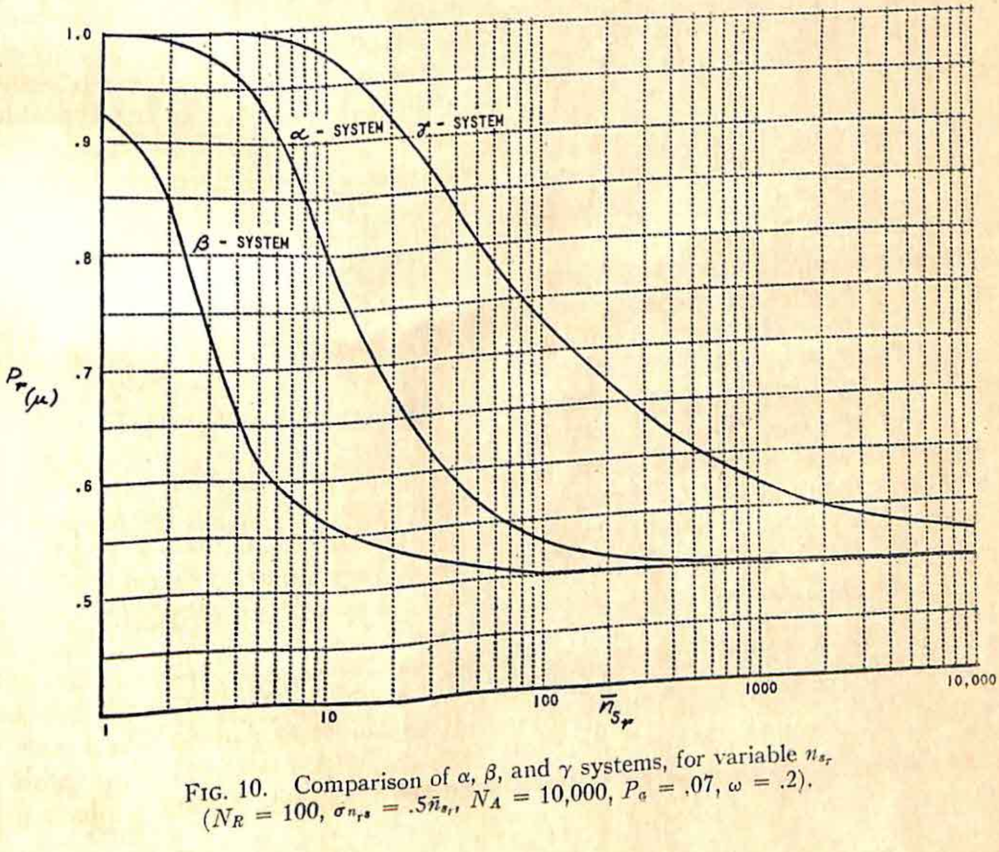


FIG. 10. Comparison of α, β, and γ systems, for variable n_{s_r} ($N_R = 100$, $\sigma_{n_{rs}} = .5\bar{n}_{s_r}$, $N_A = 10{,}000$, $P_a = .07$, $\omega = .2$).

vironment" assumptions with a model for a "differentiated environment," in which several distinguishable classes of stimuli are present (such as squares, circles, and triangles, or the letters of the alphabet). If we then design an experiment in which the stimuli associated to each response are drawn from a different class, then the learning curves of the perceptron are drastically altered. The most important difference is that the constant c_1 (the coefficient of n_{s_r} in the numerator of Z) is no longer equal to zero, so that Equation 4 now has a nonrandom asymptote. Moreover, in the form for P_g (the probability of correct generalization), where $c_2 = 0$, the quantity Z remains greater than zero, and P_g actually approaches the same asymptote as P_r. Thus the equation for the perceptron's performance after infinite experience with each class of stimuli is identical for P_r and P_g:

$$P_{r\infty} = P_{g\infty} = [1 - (1 - P_a)^{N_e}] \times \phi\left(\frac{c_1}{\sqrt{c_3}}\right) \quad (9)$$

This means that *in the limit it makes no difference whether the perceptron has seen a particular test stimulus before or not; if the stimuli are drawn from a differentiated environment, the performance will be equally good in either case.*

In order to evaluate the performance of the system in a differentiated environment, it is necessary to define the quantity $P_{c\alpha\beta}$. This quantity is interpreted as the expected value of P_c between pairs of stimuli drawn at random from classes α and β. In particular, P_{c11} is the expected value of P_c between members of the same class, and P_{c12} is the expected value of P_c between an S_1 stimulus drawn from Class 1 and an S_2 stimulus drawn from Class 2. P_{c1x} is the expected value of P_c between members of Class 1 and stimuli drawn at random from all other classes in the environment.

If $P_{c11} > P_a > P_{c12}$, the limiting performance of the perceptron ($P_{g\infty}$) will be better than chance, and learning of some response, R_1, as the proper "generalization response" for members of Class 1 should eventually occur. If the above inequality is not met, then improvement over chance performance may not occur, and the Class 2 response is likely to occur instead. It can be shown (**15**) that for most simple geometrical forms, which we ordinarily regard as "similar," the required inequality can be met, if the parameters of the system are properly chosen.

The equation for P_r, for the sum-discriminating version of an α-perceptron, in a differentiated environment where n_{s_r} is fixed for all responses, will have the following expressions for the four coefficients:

$$\left.\begin{aligned}
c_1 &= P_a N_e (P_{c11} - P_{c12}) \\
c_2 &= P_a N_e (1 - P_{c11}) \\
c_3 &= \sum_{r=1,2} P_a (1 - P_a) N_e \\
&\quad \times [P_{c\,1r}^{\,2} + \sigma_3^2(P_{c1r}) \\
&\quad + \sigma_j^2(P_{c1r}) + (\omega N_R - 1)^2 \\
&\quad \times (P_{c1x} + \sigma_s^2(P_{c1x}) \\
&\quad + \sigma_j^2(P_{crx})) + 2(\omega N_R - 1) \\
&\quad (P_{c1r} P_{c1x})] + P_a^2 N_c^2 \\
&\quad \times [\sigma_s^2(P_{c1r}) + (\omega N_R - 1)^2 \\
&\quad \times \sigma_s^2(P_{c1x}) + 2(\omega N_R - 1)\epsilon] \\
c_4 &= \sum_{r=1,2} P_a N_c [P_{c1r} - P_{c1r}^{\;2} \\
&\quad - \sigma_s^2(P_{c1r}) - \sigma_j^2(P_{c1r}) \\
&\quad + (\omega N_R - 1)(P_{c1x} - P_{c1x}^{\;2} \\
&\qquad - \sigma_j^2(P_{c1x}))]
\end{aligned}\right\} \quad (10)$$

where

$\sigma_s^2(P_{c1r})$ and $\sigma_s^2(P_{c1x})$ represent the variance of P_{c1r} and P_{c1x} measured over the set of possible test stimuli, S_t, and

$\sigma_j^2(P_{c1r})$ and $\sigma_j^2(P_{c1x})$ represent the variance of P_{c1r} and P_{c1x} measured over the set of all A-units, a_j.

ϵ = covariance of $P_{c1r}P_{c1x}$, which is assumed to be negligible.

The variances which appear in these expressions have not yielded, thus far, to a precise analysis, and can be treated as empirical variables to be determined for the classes of stimuli in question. If the sigma is set equal to half the expected value of the variable, in each case, a conservative estimate can be obtained. When the stimuli of a given class are all of the same shape, and uniformly distributed over the retina, the subscript s variances are equal to zero. $P_{g(\Sigma)}$ will be represented by the same set of coefficients, except for c_2, which is equal to zero, as usual.

For the mean-discriminating system, the coefficients are:

$$\left.\begin{aligned}
c_1&=(P_{c11}-P_{c12})\\
c_2&=(1-P_{c11})\\
c_3&=\sum_{r=1,2}\left[\frac{1}{P_a(N_e-1)}-\frac{1}{N_e-1}\right]\\
&\quad\times[\sigma_j^2(P_{c1r})+(\omega N_R-1)^2\\
&\quad\times\sigma_j^2(P_{c1x})]+[\sigma_s^2(P_{c1r})\\
&\qquad+(\omega N_R-1)^2\sigma_s^2(P_{c1x})]\\
c_4&=\sum_{r=1,2}\frac{1}{P_aN_e}[P_{c1r}-P_{c1r}^2\\
&\quad-\sigma_s^2(P_{c1r})-\sigma_j^2(P_{c1r})\\
&\quad+(\omega N_R-1)(P_{c1x}-P_{c1x}^2\\
&\qquad-\sigma_s^2(P_{c1x})-\sigma_j^2(P_{c1x}))]
\end{aligned}\right\}\quad(11)$$

Some covariance terms, which are considered negligible, have been omitted here.

A set of typical learning curves for the differentiated environment model is shown in Fig. 11, for the mean-discriminating system. The parameters are based on measurements for a

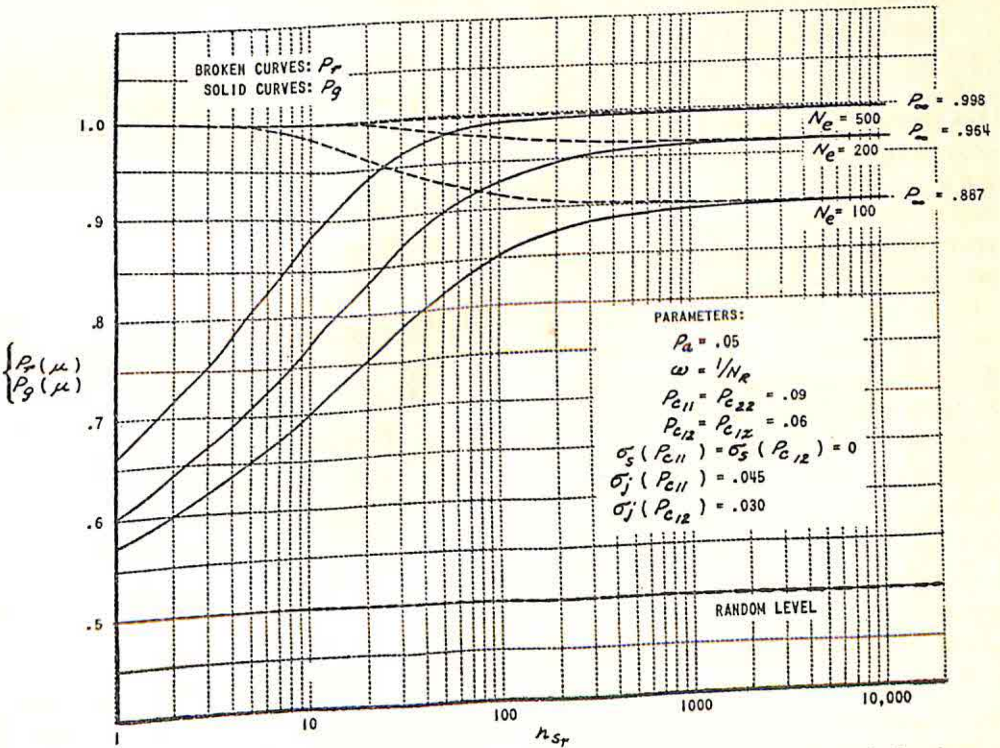


FIG. 11. P_r and P_g as function of n_{s_r}. Parameters based on square-circle discrimination.

square-circle discrimination problem. Note that the curves for P_r and P_g both approach the same asymptotes, as predicted. The values of these asymptotes can be obtained by substituting the proper coefficients in Equation 9. As the number of association cells in the system increases, the asymptotic learning limit rapidly approaches unity, so that for a system of several thousand cells, the errors in performance should be negligible on a problem as simple as the one illustrated here.

As the number of responses in the system increases, the performance becomes progressively poorer, if every response is made mutually exclusive of all alternatives. One method of avoiding this deterioration (described in detail in Rosenblatt, **15**) is through the binary coding of responses. In this case, instead of representing 100 different stimulus patterns by 100 distinct, mutually exclusive responses, a limited number of discriminating features is found, each of which can be independently recognized as being present or absent, and consequently can be represented by a single pair of mutually exclusive responses. Given an ideal set of binary characteristics (such as dark, light; tall, short; straight, curved; etc.), 100 stimulus classes could be distinguished by the proper configuration of only seven response pairs. In a further modification of the system, a single response is capable of denoting by its activity or inactivity the presence or absence of each binary characteristic. The efficiency of such coding depends on the number of independently recognizable "earmarks" that can be found to differentiate stimuli. If the stimulus can be identified only in its entirety and is not amenable to such analysis, then ultimately a separate binary response pair, or bit, is required to denote the presence or absence of each stimulus class (e.g., "dog" or "not dog"), and nothing has been gained over a system where all responses are mutually exclusive.

BIVALENT SYSTEMS

In all of the systems analyzed up to this point, the increments of value gained by an active A-unit, as a result of reinforcement or experience, have always been *positive*, in the sense that an active unit has always *gained* in its power to activate the responses to which it is connected. In the gamma-system, it is true that some units lose value, but these are always the *inactive units*, the active ones gaining in proportion to their rate of activity. In a *bivalent system*, two types of reinforcement are possible (positive and negative), and an active unit may either gain or lose in value, depending on the momentary state of affairs in the system. If the positive and negative reinforcement can be controlled by the application of external stimuli, they become essentially equivalent to "reward" and "punishment," and can be used in this sense by the experimenter. Under these conditions, a perceptron appears to be capable of trial-and-error learning. A bivalent system need not necessarily involve the application of reward and punishment, however. If a binary-coded response system is so organized that there is a single response or response-pair to represent each "bit," or stimulus characteristic that is learned, with positive feedback to its own source-set if the response is "on," and negative feedback (in the sense that active A-units will lose rather than gain in value) if the response is "off," then the system is still bivalent in its characteristics. Such a bivalent system is particularly efficient in reducing some of the bias effects (preference for the wrong response due to greater size or frequency of its associated stimuli) which plague the alternative systems.

Several forms of bivalent systems have been considered (**15**, Chap. VII). The most efficient of these has the following logical characteristics.

If the system is under a state of positive reinforcement, then a positive ΔV is added to the values of all active A-units in the source-sets of "on" responses, while a negative ΔV is added to the active units in the source-sets of "off" responses. If the system is currently under negative reinforcement, then a negative ΔV is added to all active units in the source-set of an "on" response, and a positive ΔV is added to active units in an "off" source-set. If the source-sets are disjunct (which is essential for this system to work properly), the equation for a bivalent γ-system has the same coefficients as the monovalent α-system, for the μ-case (Equation 11).

The performance curves for this system are shown in Fig. 12, where the asymptotic generalization probability attainable by the system is plotted for the same stimulus parameters that were used in Fig. 11. This is the probability that all bits in an n-bit response pattern will be correct. Clearly, if a majority of correct responses is sufficient to identify a stimulus correctly, the performance will be better than these curves indicate.

In a form of bivalent system which utilizes more plausible biological assumptions, A-units may be either excitatory or inhibitory in their effect on connected responses. A positive ΔV in this system corresponds to the incrementing of an excitatory unit, while a negative ΔV corresponds to the incrementing of an inhibitory unit. Such a system performs similarly to the one considered above, but can be shown to be less efficient.

Bivalent systems similar to those illustrated in Fig. 12 have been simulated in detail in a series of experiments with the IBM 704 computer at the Cornell Aeronautical Laboratory. The results have borne out the theory in all of its main predictions, and will be reported separately at a later time.

IMPROVED PERCEPTRONS AND SPONTANEOUS ORGANIZATION

The quantitative analysis of perceptron performance in the preceding sections has omitted any consideration of *time* as a stimulus dimension. A perceptron which has no capability for temporal pattern recognition is referred to as a "momentary stimulus perceptron." It can be shown (**15**) that the same principles of statistical separability will permit the perceptron to distinguish velocities, sound sequences, etc., provided the stimuli leave some temporarily persistent trace, such as an altered threshold,

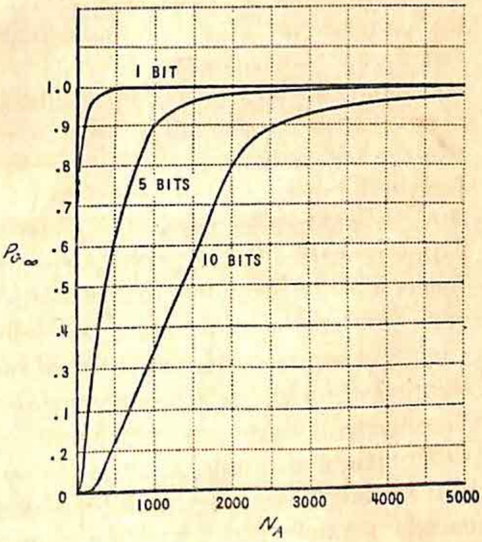


FIG. 12. P_{G_∞} for a bivalent binary system (same parameters as Fig. 11).

which causes the activity in the A-system at time t to depend to some degree on the activity at time $t - 1$.

It has also been assumed that the origin points of A-units are completely random. It can be shown that by a suitable organization of origin points, in which the spatial distribution is constrained (as in the projection area origins shown in Fig. 1), the A-units will become particularly sensitive to the location of contours, and performance will be improved.

In a recent development, which we hope to report in detail in the near future, it has been proven that if the values of the A-units are allowed to decay at a rate proportional to their magnitude, a striking new property emerges: the perceptron becomes capable of "spontaneous" concept formation. That is to say, if the system is exposed to a random series of stimuli from two "dissimilar" classes, and all of its responses are automatically reinforced without any regard to whether they are "right" or "wrong," the system will tend towards a stable terminal condition in which (for each binary response) the response will be "1" for members of one stimulus class, and "0" for members of the other class; i.e., the perceptron will spontaneously recognize the difference between the two classes. This phenomenon has been successfully demonstrated in simulation experiments, with the 704 computer.

A perceptron, even with a single logical level of A-units and response units, can be shown to have a number of interesting properties in the field of selective recall and selective attention. These properties generally depend on the intersection of the source sets for different responses, and are elsewhere discussed in detail (**15**). By combining audio and photo inputs, it is possible to associate sounds, or auditory "names" to visual objects, and to get the perceptron to perform such selective responses as are designated by the command "Name the object on the left," or "Name the color of this stimulus."

The question may well be raised at this point of where the perceptron's capabilities actually stop. We have seen that the system described is sufficient for pattern recognition, associative learning, and such cognitive sets as are necessary for selective attention and selective recall. The system appears to be potentially capable of temporal pattern recognition, as well as spatial recognition, involving any sensory modality or combination of modalities. It can be shown that with proper reinforcement it will be capable of trial-and-error learning, and can learn to emit ordered sequences of responses, provided its own responses are fed back through sensory channels.

Does this mean that the perceptron is capable, without further modification in principle, of such higher order functions as are involved in human speech, communication, and thinking? Actually, the limit of the perceptron's capabilities seems to lie in the area of relative judgment, and the abstraction of relationships. In its "symbolic behavior," the perceptron shows some striking similarities to Goldstein's brain-damaged patients (**5**). Responses to definite, concrete stimuli can be learned, even when the proper response calls for the recognition of a number of simultaneous qualifying conditions (such as naming the color if the stimulus is on the left, the shape if it is on the right). As soon as the response calls for the recognition of a relationship between stimuli (such as "Name the object left of the square." or "Indicate the pattern that appeared before the circle."), however, the problem generally becomes excessively difficult for the perceptron. Statistical separability alone does not provide a sufficient basis for higher order abstraction. Some system, more advanced in principle than the perceptron, seems to be required at this point.

CONCLUSIONS AND EVALUATION

The main conclusions of the theoretical study of the perceptron can be summarized as follows:

1. In an environment of random stimuli, a system consisting of randomly connected units, subject to the parametric constraints discussed above, can learn to associate specific responses to specific stimuli. Even if many stimuli are associated to each response, they can still be recognized with a better-than-chance probability, although they may resemble one another closely and may activate many of the same sensory inputs to the system.
2. In such an "ideal environment," the probability of a correct response diminishes towards its original random level as the number of stimuli learned increases.
3. In such an environment, no basis for generalization exists.
4. In a "differentiated environment," where each response is associated to a distinct class of mutually correlated, or "similar" stimuli, the probability that a learned association of some specific stimulus will be correctly retained typically approaches a better-than-chance asymptote as the number of stimuli learned by the system increases. This asymptote can be made arbitrarily close to unity by increasing the number of association cells in the system.
5. In the differentiated environment, the probability that a stimulus *which has not been seen before* will be correctly recognized and associated to its appropriate class (the probability of correct generalization) approaches the same asymptote as the probability of a correct response to a previously reinforced stimulus. This asymptote will be better than chance if the inequality $P_{c12} < P_a < P_{c11}$ is met, for the stimulus classes in question.
6. The performance of the system can be improved by the use of a contour-sensitive projection area, and by the use of a binary response system, in which each response, or "bit," corresponds to some independent feature or attribute of the stimulus.
7. Trial-and-error learning is possible in bivalent reinforcement systems.
8. Temporal organizations of both stimulus patterns and responses can be learned by a system which uses only an extension of the original principles of statistical separability, without introducing any major complications in the organization of the system.
9. The memory of the perceptron is *distributed*, in the sense that any association may make use of a large proportion of the cells in the system, and the removal of a portion of the association system would not have an appreciable effect on the performance of any one discrimination or association, but would begin to show up as a general deficit in *all* learned associations.
10. Simple cognitive sets, selective recall, and spontaneous recognition of the classes present in a given environment are possible. The recognition of relationships in space and time, however, seems to represent a limit to the perceptron's ability to form cognitive abstractions.

Psychologists, and learning theorists in particular, may now ask: "What has the present theory accomplished, beyond what has already been done in the quantitative theories of Hull, Bush and Mosteller, etc., or physiological theories such as Hebb's?" The present theory is still too primitive, of course, to be considered as a full-fledged rival of existing theories of human learning. Nonetheless, as a first approximation, its chief accomplishment might be stated as follows:

For a given mode of organization (α, β, or γ; Σ or μ; monovalent or bivalent) the fundamental phenomena of *learning, perceptual discrimination, and generalization can be predicted entirely from six basic physical parameters*, namely:

x: the number of excitatory connections per A-unit,

y: the number of inhibitory connections per A-unit,

θ: the expected threshold of an A-unit,

ω: the proportion of R-units to which an A-unit is connected,

N_A: the number of A-units in the system, and

N_R: the number of R-units in the system.

N_s (the number of sensory units) becomes important if it is very small. It is assumed that the system begins with all units in a uniform state of value; otherwise the initial value distribution would also be required. *Each of the above parameters is a clearly defined physical variable, which is measurable in its own right, independently of the behavioral and perceptual phenomena which we are trying to predict.*

As a direct consequence of its foundation on physical variables, the present system goes far beyond existing learning and behavior theories in three main points: parsimony, verifiability, and explanatory power and generality. Let us consider each of these points in turn.

1. *Parsimony.* Essentially all of the basic variables and laws used in this system are already present in the structure of physical and biological science, so that we have found it necessary to postulate only one hypothetical variable (or construct) which we have called V, the "value" of an association cell; this is a variable which must conform to certain functional characteristics which can clearly be stated, and which is assumed to have a potentially measurable physical correlate.

2. *Verifiability.* Previous quantitative learning theories, apparently without exception, have had one important characteristic in common: they have all been based on measurements of *behavior*, in specified situations, using these measurements (after theoretical manipulation) to predict *behavior* in other situations. Such a procedure, in the last analysis, amounts to a process of curve fitting and extrapolation, in the hope that the constants which describe one set of curves will hold good for other curves in other situations. While such extrapolation is not necessarily circular, in the strict sense, it shares many of the logical difficulties of circularity, particularly when used as an "explanation" of behavior. Such extrapolation is difficult to justify in a new situation, and it has been shown that if the basic constants and parameters are to be derived anew for any situation in which they break down empirically (such as change from white rats to humans), then the basic "theory" is essentially irrefutable, just as any successful curve-fitting equation is irrefutable. It has, in fact, been widely conceded by psychologists that there is little point in trying to "disprove" any of the major learning theories in use today, since by extension, or a change in parameters, they

have all proved capable of adapting to any specific empirical data. This is epitomized in the increasingly common attitude that a choice of theoretical model is mostly a matter of personal aesthetic preference or prejudice, each scientist being entitled to a favorite model of his own. In considering this approach, one is reminded of a remark attributed to Kistiakowsky, that "given seven parameters, I could fit an elephant." This is clearly *not* the case with a system in which the independent variables, or parameters, can be measured *independently* of the predicted behavior. In such a system, it is not possible to "force" a fit to empirical data, if the parameters in current use should lead to improper results. In the current theory, a failure to fit a curve in a new situation would be a clear indication that either the theory or the empirical measurements are wrong. Consequently, if such a theory *does* hold up for repeated tests, we can be considerably more confident of its validity and of its generality than in the case of a theory which must be hand-tailored to meet each situation.

3. *Explanatory power and generality.* The present theory, being derived from basic physical variables, is not specific to any one organism or learning situation. It can be generalized in principle to cover any form of behavior in any system for which the physical parameters are known. A theory of learning, constructed on these foundations, should be considerably more powerful than any which has previously been proposed. It would not only tell us what behavior might occur in any known organism, but would permit the *synthesis* of behaving systems, to meet special requirements. Other learning theories tend to become increasingly qualitative as they are generalized. Thus a set of equations describing the effects of reward on T-maze learning in a white rat reduces simply to a statement that rewarded behavior tends to occur with increasing probability, when we attempt to generalize it from any species and any situation. The theory which has been presented here loses none of its precision through generality.

The theory proposed by Donald Hebb (7) attempts to avoid these difficulties of behavior-based models by showing how psychological functioning might be derived from neurophysiological theory. In his attempt to achieve this, Hebb's philosophy of approach seems close to our own, and his work has been a source of inspiration for much of what has been proposed here. Hebb, however, has never actually achieved a model by which behavior (or any psychological data) can be *predicted* from the physiological system. His physiology is more a suggestion as to the *sort* of organic substrate which might underlie behavior, and an attempt to show the plausibility of a bridge between biophysics and psychology.

The present theory represents the first actual completion of such a bridge. Through the use of the equations in the preceding sections, it is possible to predict learning curves from neurological variables, and likewise, to predict neurological variables from learning curves. How well this bridge stands up to repeated crossings remains to be seen. In the meantime, the theory reported here clearly demonstrates the feasibility and fruitfulness of a quantitative statistical approach to the organization of cognitive systems. By the study of systems such as the perceptron, it is hoped that those fundamental laws of organization which are common to all information handling systems, machines and men included, may eventually be understood.

REFERENCES

1. ASHBY, W. R. *Design for a brain.* New York: Wiley, 1952.
2. CULBERTSON, J. T. *Consciousness and behavior.* Dubuque, Iowa: Wm. C. Brown, 1950.
3. CULBERTSON, J. T. Some uneconomical robots. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton Univer. Press, 1956. Pp. 99–116.
4. ECCLES, J. C. *The neurophysiological basis of mind.* Oxford: Clarendon, 1953.
5. GOLDSTEIN, K. *Human nature in the light of psychopathology.* Cambridge: Harvard Univer. Press, 1940.
6. HAYEK, F. A. *The sensory order.* Chicago: Univer. Chicago Press, 1952.
7. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
8. KLEENE, S. C. Representation of events in nerve nets and finite automata. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton Univer. Press, 1956. Pp. 3–41.
9. KÖHLER, W. Relational determination in perception. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 200–243.
10. MCCULLOCH, W. S. Why the mind is in the head. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 42–111.
11. MCCULLOCH, W. S., & PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bull. math. Biophysics*, 1943, **5**, 115–133.
12. MILNER, P. M. The cell assembly: Mark II. *Psychol. Rev.*, 1957, **64**, 242–252.
13. MINSKY, M. L. Some universal elements for finite automata. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton Univer. Press, 1956. Pp. 117–128.
14. RASHEVSKY, N. *Mathematical biophysics.* Chicago: Univer. Chicago Press, 1938.
15. ROSENBLATT, F. *The perceptron: A theory of statistical separability in cognitive systems.* Buffalo: Cornell Aeronautical Laboratory, Inc. Rep. No. VG-1196-G-1, 1958.
16. UTTLEY, A. M. Conditional probability machines and conditioned reflexes. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton Univer. Press, 1956. Pp. 253–275.
17. VON NEUMANN, J. The general and logical theory of automata. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 1–41.
18. VON NEUMANN, J. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton Univer. Press, 1956. Pp. 43–98.

(Received April 23, 1958)

BUREAU OF EDNL. & PSYL.
LIBRARY